

Db 14 LLAALCAAGALEEKVCOGTSNRLTQLTFFDHFLSLQRMFNNECVILGNLEITYVQRN 73  
Qy 69 ASLSFLQDIQEOGVYVLIHNOVQVPLORLRIVRGTQTLFEDNYALAVLDNGDPLNNTTP 128  
Db 74 YDLSFLKTTQEVAGYVLIHNTVERIPLENLQIRGNALYENTYALAVLSN-----124  
Qy 129 VTGASPGGLRELQSLRLTILKGGVLIQRNPOLCYODTILKWDIFHKNQALATLIDTNR 188  
Db 125 -YGTNKTGLURELPMRNLQILIGAVRFSNNPLCNMETIOWRDI--QDVFLSNMSMDVQR 182  
Qy 189 S-RACHPCSPMCKGRCWGESSEDCQSLTRTVCAGGCA-RCKGPLPTDCHEQCAAGCTG 246  
Db 183 HLTGCPKCDPSCPNGSCWGRGENCEKLTIIICAQOCSSRCGRSPSDCHNQCAAGCTG 242  
Qy 247 PKHSOCLACLHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLST 306  
Db 243 PRESCLVCHRDEATCKDTCPLMLYNPTTYQMDVNPGEKYSGFATCVKCPRYVVT 302  
Qy 307 DVGSCITLVCPLHNOEVAEDGTORCEKCKSPCARVCYGLGMQYIKANSFIGITELE-PA 365  
Db 303 DHGSCVRACPDYEV-EEDGVSCKKCDGCRKVCNGIGIGEFK-DTISINATNIKFK 360  
Qy 366 GCKKIFGSLAFIPESFGDGPASNTAPLOPEQLQVFPETLEITGYLYISAWPDSLPLSVF 425  
Db 361 YCTAISGDLHLPLVAFKGSFTRTPPLDPRELEILKTVKEITGFLLIQAWPENWTDLHAF 420  
Qy 426 QNLQVIRGRILNGAYSLTLOGLGISWIGLSRLSRLGSLGLAIHNNTHLCFVHTVPWDL 485  
Db 421 ENLEIRGRTHQGOFSLAVVGLNITSLGRLSRLKEISDGDVITISGNRNLCYANTINWKL 480  
Qy 486 FRNHQALLHTANRDEDECVGGLACHOLCARGHCGPGTCTOCVNCQSQFLRGOEVECR 545  
Db 481 FGTPNQKTKIMNRAEKKCKATNHCNPLCSGEGMGPEPTDCVSCQNVSRGREGVDKCN 540  
Qy 546 VLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARCPGVK 605  
Db 541 ILEGEPREFVENSECICQCHPECLPQTMNITCTGRGPDNCIKAHYVDGPHCVKTCPSGIM 600  
Qy 606 PDLSTMPYTWKFPDEGACQCPINCTHSCVDLDDKGCAPAEQASRP-LTSIVSAVY-GILL 663  
Db 601 GENNTL-VNKFADANNVCHLCHANCTYGCAGPLAGC--QOPEGPKIPSIATGIVGGLIF 657  
Qy 664 VVLGVVFGILI--QYIKANS--KFIGITEL--PLTPSGAMPNQAQWRILKTELKVK 716  
Db 658 IVVVALGIGLFWRRQLVKRTLRLQLRELVLELTPSGEAPNQAHLRLKTEFFKIK 717  
Qy 717 VLGSAGFTVYKGIWIPGENVKIPIVAKVLRENTSPKANKILDEAYVMAGVGPVYSR 776  
Db 718 VLGSAGFTVYKGLWIPGEKVKIPIVAKELREATSPKANKILDEAYVMASVDNPHVCR 777  
Qy 777 LIGICLTSTVQLVTLQMPYGCILLDHVRENRLGSDQLLNNWCQIAKGMVSYLEDVRLVHR 836  
Db 778 LIGICLTSTVQLITQMPYGCILLDVREHKDNIGSQYLLNWCQIAKGMVSYLEDVRLVHR 837  
Qy 837 DLAARNVLKSPNHVKITDFGLARLLDDETEYHADGGKVPKKNWALSILRRRTHOSD 896  
Db 838 DLAARNVLKTPQHKITDFGLAKULGAEKEHYHAEGGKVPKKNWALSILRIYTHOSD 897  
Qy 897 VWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPCTIDVYIMVWKMWIDSE 956  
Db 898 VWSYGVTVWELMTFGSKPYDGIPIASEISSILEKGERLPQPPCTIDVYIMVWKMWIDAD 957  
Qy 957 CRPRFELVBSFERNARPPQRFVWIQ-NEDLGPASPLDSTFYRSLLEDDDDMGDLVDABEY 1015  
Db 958 SRPKFELILEFSKWARDPQRYLVIQGDERMHLPSPTDSNFYRLMEBEDMEDVVDADBY 1017  
Qy 1016 LVPOQGFCDPAPCAGGMVHRHRSSTSRSGGDLTLGLEPSEEAEPRLAPSEGAGS 1075  
Db 1018 LIPOQGF-----NSPST-----SRTPLSLLSLANS 1043  
Qy 1076 DVFDGDLGMGAAGKGLQSLPHTDPSPLQRYSEDPVPLPSET--DGAVAPLTCSPQPEVYN 1133  
Db 1044 N-----SSTVACINRNGSCRVEDAFQRYSDPTSLVDNIDDTFL-----PVPEVIN 1093

Qy 1134 QPDVRRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVVVKOVFAFGGAVENPEYL-TPQ 1192  
Db 1094 Q-SVPRKPAQSVQNPVYHNPQLHP-----APGRDLHYQN--PHSNAVSNPEYLNTAQ 1142  
Qy 1193 GCAGAPQHPPPAFSPAFDNLYYWQ-----DP-----PERCAPSTFKGTPTAE 1236  
Db 1143 -----PTCLSSGFDSSALWIKQSHQMSLDNPDYQODFPFPEAKNGIFKG-PTAE 1192  
Qy 1237 NPEYLGLDVP 1246  
Db 1193 NAEYLRVAPP 1202

## RESULT 3

Q9EP98 PRELIMINARY; PRT; 1210 AA.  
AC Q9EP98;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Epidermal growth factor receptor isoform 1.  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;  
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
RA Maible N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and  
mouse alternative Egr transcripts encoding truncated receptor  
isoforms.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J;  
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
RA Maible N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and  
mouse alternative Egr transcripts encoding truncated receptor  
isoforms.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF2753366; AAG28045.1;  
DR EMBL; AF2753364; AAG28045.1; JOINED.  
DR EMBL; AF2753365; AAG28045.1; JOINED.  
DR EMBL; AF2753367; AAG24386.1;  
DR HSSP; F11362; IFGK.  
DR MGD; MGI:95294; Egr.  
DR InterPro; IPR000345; CytC heme bind.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD0000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.

[illegible]

SQ	SEQUENCE	1165 AA; 129614 MW; 7F7EE38D8771A74E CRC64;
Query Match	39.0%; Score 2646; DB 13; Length 1165;	
Best Local Similarity	44.7%; Pred. No. 8.7e-190;	
Matches	570; Conservative 163; Mismatches 388; Indels 154; Gaps 32;	
QY	1 MELAALCRWGLLLALLPPG-AAST-----QVCTGTDMKRLRPASPETHLDMRLHLYQGCQV 55	
DB	4 LELLEL-----LLLLLSIGRCCSDPRKVCQQTSNQTM-----LDNHYLXMKMKYSGCV 56	
QY	56 VQGNLELTPTNASTLSFLODIQVGVLIHNOVRQVPLRLIRIVRGTLQFEDNYALA 115	
DB	57 VLENLEITYQENQDLPSLQISQVGVLIAMNEVSTIPLVNLRLIRGNLYEGNFTLL 116	
QY	116 VLNDGDPPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNQPCYQDITLWKDIFHK 175	
DB	117 VMSNYQK-PPSSP-DVYQVGLKQLQSLNLTETLLSGGVKVSHPNLLCNVETINWMDIVDK 173	
QY	176 NNQALTLIDTNRSRACHPCSPMCKSRGWGESSEDCQSLTRTVCAGCC-ARCKGPLPTD 234	
DB	174 TSNPTMNLIPHAFERQCKCDPCGVNGSCWAPGFGHCQKFTKLLCABQCNRRCRGRPID 233	
QY	235 CCEOCAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFAS 294	
DB	234 CCEHCAGCTGPRATDCLACRDNDGCTCKTCTPPPKIYDIYSHQVVDNPNIKYTFGA 293	
QY	295 CVTACPNYLSLTDVGSCTVCLPLHNOBVTAEADGTQRCCKSKPCARVCYGLGM---QYI 350	
DB	294 CVKCEPSNVVTE-GACVRSKCSAGLEVD-ENGRSKCKPCDGVCPKCDGIGIGLSNTI 351	
QY	351 KANSKFIGITELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLQVFEETEEITGYL 410	
DB	352 AVNSTWIG-----SFSNCTKINGDIILNRNSFEQDPHYKIGMPDPEHLWNLTTVKEITGYL 407	
QY	411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYS-LTLOGLISWLGSLRELSGSLALIH 469	
DB	408 VIMWPMNTSLSVFQNLQVIRGRILHNGAYS-LTLOGLISWLGSLRELSGSLALIH 467	
QY	470 HNTLFCVHTVPMDQLFNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPQTQCV 529	
DB	468 NTPQLRYASTINMRLFRSEDSQSTEYDART-----ENQTCNECSEDGCGWGPPTMCV 520	
QY	530 NCQOFIRGCQVCECRVQLGLPREYVNRHCLPCHPECPQONGSVTCFGEAPAOQVACAH 589	
DB	521 SCLHVRGRGRCVASCNLLQGEPREAQVQDGRVCQHQECLVQDTSLTTCYGPANCSKCAH 580	
QY	590 YKDPFPCVAPCSGVKPDLSYMPITWPKFDEEGACQPCINCTHSCVDLDDKQCPAEORAS 649	
DB	581 FQGPQCIPRPHGMGLDGDUTL-IWKYADKMGQCPCHQNCQCGSGPGLSGCRGD-IVS 638	
QY	650 PLTSIVSAVVGILLVVLGVVFGILIQYIKANSK-----FIGITEL--PLTPSGAMPNQA 702	
DB	639 HSSLAVGLVGLLITVIVALLIVLLRRRRIKRRTIRRLLOEKELVEPLTPSGAPNQA 698	
QY	703 QMRILKETELRKVKVLGSGAFYVYKGINIPDGENVKIPVAIKVLRNTPSKANKETLDE 762	
DB	699 FLRLKETEFKQDVLGSGAFYVYKGLMNPDPGENIRIPVAIKVLRNTPSKANKETLDE 758	
QY	763 AVYVAGVGSYVSRLLGICLTSVQLVTLQMPYGCLLDHRNENRGLSGODLNNWCQOIA 822	
DB	759 AVYVASVDHPHVCRLGICLTSVQLVTLQMPYGCLLDHRNENRGLSGODLNNWCQOIA 818	
QY	823 KGSYLEVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVPIKQWMA 882	
DB	819 KGNVYLEERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADKEYQAHGKVPKQWMA 878	
QY	883 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 942	
DB	879 LESILQWYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPPICTIE 938	
QY	943 VYMIMVKCMWIDSECRPRELVESESRWADPQRFVWQNEQDLGPASPLDSTFYSLLE 1002	
DB	939 VTMILKCMWIDSPSRPRELVESESRWADPQRFVWQNEQDLGPASPLDSTFYSLLE 995	

QY	1003 DDMDGLVDABEYLVPOQGFPCPDPAQAGCMVHHRSSSTRSGGDLTLGLPESEEA 1062	
DB	996 SDD--DVVDADEYLL-----RYKRIN-RQS----- 1018	
QY	1063 PRSPAPSEAGSAGSVDFDGLGMGAAGKGLQSLTPHDPGLQRYSEDPTV-PLPSETDGYVA 1121	
DB	1019 --EPCIPPNHG-----PVRENSIALRYISDPTQNALEKDLGDH-- 1054	
QY	1122 PLTCSPOPEYVNOBVPQP-----PSPRE-----GPLP-AARPAGATLERAKTILSPG 1168	
DB	1055 -----EYVNOGSETSSRLSDIYNPNVEDLTGWSVPSLSQEAETNFSRPEYLNTN 1106	
QY	1169 KNGVVKDVAFAGGAVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPST 1228	
DB	1107 QNSL---PLVSSGSMDDPDY---QAG-----YQAAE-----LPQTGALTGN 1141	
QY	1229 PKGTPTAENPEYLGL 1243	
DB	1142 GMFLPAAENLEYLGL 1156	
RESULT 5		
Q9W6F6		
ID	Q9W6F6 PRELIMINARY; PRT; 1137 AA.	
AC	Q9W6F6;	
DT	01-NOV-1999 (T-EMBLrel. 12, Created)	
DT	01-NOV-1999 (T-EMBLrel. 12, Last sequence update)	
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)	
DE	Receptor tyrosine kinase (fragment).	
GN	ERBB4.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OC	NCBI_TaxID=9031;	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=HINDRAIN;	
RX	MEDLINE=99263203; PubMed=10328884;	
RA	Dixon M., Lumsden A.;	
RT	"Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain."	
RL	Mol. Cell. Neurosci. 13:237-258 (1999).	
DR	EMBL; AI2121963; AAD31764.1; --	
DR	HSSP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L_domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001368; TNFR_c6.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 1.	
DR	Pfam; PF02757; YLP_2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.	
KW	Kinase; Tyrosine-protein kinase.	
FT	NON_TER 1	
SQ	SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;	
Query Match	39.0%; Score 2643.5; DB 13; Length 1137;	
Best Local Similarity	46.2%; Pred. No. 1.3e-189;	
Matches	528; Conservative 173; Mismatches 350; Indels 91; Gaps 28;	
QY	161 LCYQDTILWKDIPKHNQALTLIDTNRSRACHPCSPMCKSRGWGESSEDCQSLTRTV 220	

```
Db 3 LCFADTHHWQDIVRNPNWASFTLVPTNGSSCGRCHKSCGTG-RCMGPTENHCQTLLTKTVC 61
Qy 221 AGGC-ARCKGPLPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTF 279
Db 62 AEQCDGRGCGYGVSDCHRECAAGCGKPDTCFACMNFDSGACVTCQPTQFYVNPFTF 121
Qy 280 ESMNPBGRYTFGASCVTACPYNYLSLTDVSGCTLVCPLNHOEVTAEQDTCRCEKSKPCA 339
Db 122 QLEHNHAKYTYGAFVKKCPHNFV-VDSSCVACACPSKMEV-EENGKMKCKTCTDCLP 179
Qy 340 RVCYGLGQYIKANSKFIGITILE-FAGCKKIFGSLAPLPESFDGDPASNTAPLOEQ 398
Db 180 KACDGIQIGSL-VSAQTVDSNSIDKFNCTKINGNLIFLVTGIRHGDVPVHTIAAINPEKLN 238
Qy 399 VFETLEITGYLYISANPDSLPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISHLGRSL 458
Db 239 IFQVREITGYLNTQSPENMTDRFVSNLVTIGRALYSLSLILKQOQITSLQFQSL 298
Qy 459 RELGSLGALIHNTHLCHFVHTVPMDQLFRNPHQALLHTANPEDECVGEGGLACHOLCARG 518
Db 299 KQISAGNIYITDSNLCYVHTVNTSLFSTPSQKTVIHRNKKAKENCTADGMVCMELGSSD 358
Qy 519 HCWGPPTQVCNCSQFIRGQCBCECRVLOGLPREYVNNARHCLPCHPECQP-QNGSVTFC 577
Db 359 GCWGPDPQCLCKFRIGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTGRTG 418
Qy 578 GPEADQCVACAHYKDPFCVACRCSGVKPDLSYMPIMKFPDEEGACQPCINCHTSCVDL 637
Db 419 GPGPDHCTKCFHFGDGNVCBCKPQGGQANSF--IPKYADEDECHPCNCTQGCRCGP 476
Qy 638 DDKGC-----PAEQRASPLTSTVSAAV-GILLVVLGVVFGILIOYIKANS--- 682
Db 477 ASHDCIYYPWTRQSTLPQAHAR-TPL--IAAGVIGLFIIVIMGLTFV---YVRKSIKK 530
Qy 683 -----KFIGITEL--PLTPSGAMPNQAOMRILKTELKVKVLSGSGAGFTYKGIWIPDG 735
Db 531 KRALLREFLE-TELVEPLTPSGTAPNQAQRLKTELKRVKVLGSGAGFTYKGIWIPDEG 589
Qy 736 ENVKIPVAIKVLRNTSPKANKETLDEAYVNAVGSPPVSRLLIGLICTSTVOLATQMLPY 795
Db 590 ETVKIPVAIKLNETTGPKANVFEMDEALINWASHDHPHLVRLGVLGSLPTIQLVTLMPH 649
Qy 796 GCLLDHVRNRRGRGLSQDLNWCQIAKMSYLEYDLVLRHDLAARNVVLKSPNHVKITD 855
Db 650 GCLLDYVHEHKNIGSQDLNWCQIAKMSYLEYDLVLRHDLAARNVVLKSPNHVKITD 709
Qy 856 FGLARLLDIDETEHADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPY 915
Db 710 FGLARLLLEGDEKEYNADGGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGKPY 769
Qy 916 DGIPAREIPDLLEKGERLPQPICTIDVYVIMVVKWIMIDSECRPRFRELVSERFMRADP 975
Db 770 DGIPTREIPDLLEKGERLPQPICTIDVYVIMVVKWIMIDSECRPRFRELVSERFMRADP 829
Qy 976 QRFVVIQNEED-LGPASPLDSTFYRSLLEDDMDGLVDAEYLYVPOQGFCCDPDAPAGGM 1034
Db 830 QRYLVIVQDDRMKLPSPNDSKFFQNLDEEDLMDWDAEYLYV-PQAFNIIPPITYSTR 888
Qy 1035 VHRHRSSTRSGGDDLTGLEPSEEAAPRS--PLAP-SEGAGSDVFDGDLGMAAGLQ 1091
Db 889 IDSNNRNFVYRDGGYAAEQGY-PMPYRAPGCIIPAPVAQATAEIIFEDTCCNGTLRKQV 947
Qy 1092 SLPTHDPSPLORYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVPQPSP 1144
Db 948 ATLAKESSTORYSADPTVFIPIRVIRGELEDGDMTMRDKPTDYLNPNVEENPFVSR 1007
Qy 1145 REGPLPAA-RPAGATLERAKTSLSPKNGVVKDF-----AFGAVENPEYILTPOGGA 1195
Db 1008 KNGDLQAVDNPEYHN-----APNGQPKAEDEVNEPLYLNTFANTLENAEYL----- 1054
Qy 1196 APQHPHPPAFSPADNLYYWDQDPPERGA--PPSTFKGTPT-----AENPE 1239
```

```
Db 1055 --KNNLPEKAKAPDNPDYWNHSLUPRSTLQHPDYLOEYKTYFKQNGRIRPIVAENPE 1112
Qy 1240 YL 1241
Db 1113 YL 1114

RESULT 6
P79754 PRELIMINARY; PRT: 1328 AA.
ID P79754 AC P79754;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErBb3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorphii; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OC NCBI_TaxID=31033;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FCG.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU_3.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
```

Query Match 33.4%; Score 2262; DB 13; Length 1328;  
Best Local Similarity 39.8%; Pred. No. 7.7e-161;  
Matches 512; Conservative 162; Mismatches 413; Indels 200; Gaps 33;

```
Qy 9 WGLLALLAPP--GAASTQ---VCTGTDMLRLPASPTHLDMLRLHLYQGCVVQGNLEL 62
Db 4 WRLTLMCVASRLRASSQTQEAQVCPGTQNGLSSTGSGENQYNLKNDRYKGCIEIMGNLEI 63
Qy 63 TYLPTNASLSFLDQIEVQGVYLIHNRQVPLQRLRIVRGTOLEFEDNYALAVLDNGDP 122
Db 64 TQIESNWFDFLKTIREVTGVYLIAMNHQFPIPLQLRVRIGNSLYERRFALSFLN--- 120
Qy 123 LNNTPTVGTASPGGLRELRLSLTEILKGGVLIOGNPOLCQDTILWKDIFHKNQLALT 182
Db 121 ----YPKDQ--PSGLNQLGLMNLTEILDGGVQIINNKYLRYGPPVWYRDII--RNNDAPIE 173
Qy 183 LIDNRSRACHPCSPMKCGRCWGESSEDCQSLTRTCAGGC-ARCKGPLPTDCCHEOCA 241
Db 174 IQFNGERGVCVH---KSC-GNYCWGPGKQDQQLTKVCAPOCNDRCFTGSPRDCCHIECA 229
Qy 242 AGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMNPBGRYTFGASCVTACPY 301
Db 230 AGCKGPLDTCFACRLFNDSGACVPQCPQTLVYNKQTFQMETPNPAKYQSGICVSCQPT 289
Qy 302 NYLSTDVGSCTLVCPLNHOEVTAEQDTCR-CEKSKPCARVCYGLGQYIKANSKFIGIT 360
Db 290 HFV-VDGSSCVSVCPDKMEV--ERGSQRQCELSCGLCPKYCEGTGAE---ORQTVDSS 342
```



```
QY 361 ELE-FAGCKKIFGSLAPLPSFQDGPASNTAPLOQLOVFETLEETGYLISAWPDSL 419
Db 343 NIDSFNCTKIQGSLHFLVGTGILGDDDFKNVPPPLDAKKEVFTVREITDILNIQSWPEL 402
QY 420 PDLVSFQNLQVIRGRILHNGAYSITLQGLGIGSWGLRLSRLSGSLALIHNTLHLCFVHT 479
Db 403 NDLVSFSSLTIIQGRSLFKRFSLMVMRIPITLSLGLSLREISDGSVYISQNAHLCHYHT 462
QY 480 VPMQDLFRNPH-QALLHTANRPEDECYGEGLACHQLCARGHCGPGTQCVCNCSQFLRGQ 538
Db 463 VNMQLFRGSRVANSLSNRPMACVADGRVCDPLCSGCGWCGPGDQCLSCRNYSRHG 522
QY 539 ECVEECVLOGLPREYNARH-CLPCHPECPQNGSVTCGPDADQCVACAHYKDPFV 597
Db 523 TCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTGKASCTGPGADECMACTKFRDGYCM 582
QY 598 ARCPGSGVKPDLSPYMTWKFDEGACOPCINPCTHSCVOLDKGCAPAEQASPLTSVSA 657
Db 583 SSCPAGVN-DGEKGLIFKFNREHCEPCHQNCCTQCGSGPLGNDL---LEAARLTISGQ 638
QY 658 VVGILLVVLGVVF-----GILQIYIKANSKFI--GITELPLTPSGAMPNQA 702
Db 639 ITGIALGVPAGLIFCLVFLPLGLMYHRGLAIRKRAMRRYLESGESPEPLGP-GEKTKV 697
QY 703 QMRILKETELRVKVLGSGAFYVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDE 762
Db 698 HARILKPSDLRKIKLGSVGVFTSVKGFWPEGETVKIPVAIKTIQDSGRQFTETIDH 757
QY 763 AYMVAGVGSYPVSRLLGICLTSTVOLVTOLMPYGCILLDHVRENRLGSLQDILNWCVOIA 822
Db 758 LLSMGSLDHYIYVRLGICPGTCLQVLTQSSHSLLEHROKHTSLDQRLNWCVOIA 817
QY 823 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDPLGLARLIDIDETEHADGKVPKIMA 882
Db 818 KGMYYLEHRVHKNAARNILLKNDYQVQISDYGVADLLYDPDKVYVSETKTPKIMA 877
QY 883 LESILRRRTHQSDVMSYGVYVWELMTFGAKPYDIPAREIPDLLEKGERLPPOPICTID 942
Db 878 LESILFRYTHQSDVMSYGVYVWEMMSFGAEPYASVQBPVSPVLEKGERLSQPAICTID 937
QY 943 VYMIMVKCMIDSECPREFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1002
Db 938 VYMIMVKCMIDENIRPTFKELASDFTRMARDPRLVIRMEG-----E 981
QY 1003 DDDMGDLVDAEYLVLPQGGFCPPDPAPGAGMVHRRHRSSTSRSGGDLTLGLEPSEEA 1062
Db 982 DSGMGEFL-----RRCGER---GLLEADLEEDDEE- 1008
QY 1063 PRSPLAPSEGAGSDVFDGLGNG---AAKGLQSLPHDPSPLQ-----RYSDDPT 1109
Db 1009 -----GLGDRFATPSLPSPSWSTSPSQINSYVMVMTQLRYD--- 1044
QY 1110 VPLPSETDGYVAPLTCSPQ-ENVNO-----PDVRPQPPSPREGPL--PAA 1152
Db 1045 --FAVSGQHIGYLPMSPSVDITRQWYQSRSLSSVRLTPDRSAFRSSREAELEDGA 1102
QY 1153 RPAGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPOGGAAPQPHPPAPSPAFDNL 1212
Db 1103 QACGIFRVR-----FGSERGN-----POGG----- 1122
QY 1213 YVWDQPPRGAPPSTFKGTPTAENPE 1239
Db 1123 ---QQRKLSTAGSPSPSKTAADEDE 1146
```

## RESULT 7

```
Q9BIH9 PRELIMINARY; PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
```

## GN EGFR.

OS Anopheles gambiae (African malaria mosquito).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;  
OC Anopheles.  
OX NCBI\_TaxID=7165;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SU4;  
RA Lycett G.J.;  
RT "Cloning, expression and localisation of the Anopheles gambiae  
epidermal growth factor receptor.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
RE EMBL; AJ301655; CAC35008.1; -;  
DR HSP; P11362; IFGK.

DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 7.  
DR SMART; SM00220; STK; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 4.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW Receptor.  
FT NON\_TER 1 1

SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.6%; Score 2007.5; DB 5; Length 1433;  
Best Local Similarity 32.3%; Pred. No. 1.2e-141;

Matches 467; Conservative 198; Mismatches 389; Indels 393; Gaps 38;

```
QY 26 CTCTDKMLRLPASPETHLDMLRHLYGCGVVOGNLELTYPNALSFLQDIOEVQGVYL 85
Db 1 CIGTNGRMSVPANREYHYKRLDRYNTCTVVDGNLEITWIGNITDNLFLQHIREVTGYL 60
QY 86 IAHQVQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTTPTVTGASPGGLREL 140
Db 61 ISLYDLPOVILPRLQIRGRITTFKLNKWEAYGLFV-----SFSHMNTL 104
QY 141 QLRSLTEILKGGVLIQRNPOLCYODTILWKDI-FHKNNQLALTIDTNRSPACHPCSPMC 199
Db 105 ELPALRDILGSGVGFNNYLNCHKMSINWEEILLAPOTSQMYTFNFSRPERVCPCHPSC 164
QY 200 KGSRCNGESSEDCQSLTRTVACGGCA--RCKGPLPTDCHEQCAAGCTGPKHSCLACLH 257
Db 165 EVG-OWGEAHNCQRFSLKNCSPQCSQGRCFGPKPRECCHLFCAGGCTGTQSDCLACKN 223
QY 258 FNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACPNYVLTSDVGSCTLVCP 317
Db 224 FYDDGVCKQECPPMQIYNTPTNYFEPNPDGKYAYGATCVRKCP-EHLLKNGACVRCCKP 282
QY 318 HNOEVTAEQGTQCEKSKPCARVCYGLGMOYIKANSKFGITELEFAGCKKIFGSLAPL 377
Db 283 GKMPPQNSE-----CVPCKGVCPTCPGEGI-----VHSDNIG-----NYKCDTIEGSL 329
QY 378 PESFDGDPASNT-----APLQEQLOVFEETLEETGYLYISAWPDSLPLDSLVFQNLQ 429
Db 330 DQSFDFGQVYTNFSGPRYIKIDPDRLEVFSTVKEITGFINQAHHPNTTTLNYFRNLE 389
QY 430 VIRGRILHNGAY-SLTLOGLGISWGLRLSRLSGSLALIHNTLHLCFVHTVPMQDLFRN 488
Db 390 VVGGRLKENLFAVYIVKTSLSKLSLKRVSNGSIVILENSDLCFVEDIDMSKKS 449
```

```
Qy 489 PHQALLHTANRPEDECVEGLACHQLCARGHCWGPGPTQCVCNCSQFLRGQECVCEBRLVQ 548
Dy 450 SDHEVMQKRNARTECHEGECSEQCSKACGCKGPPQCLCKNKKYKGLCLDSCK--- 506
Qy 549 GLPREY-VNARHCLPCHPECPQNGSVTCFQPEADQCVACAHYKDPFCVACRP----- 601
Dy 507 SLPLYSVDSKTCGDCHQCKD-----FCYGNEDNCSCNMVXKDRGFCVAECPTTKHAM 561
Qy 602 -----SGVKPDLSPYMPKRPD----- 618
Dy 562 NGTCINCHKTCVGRGRPRDTIAPGCCISCDKXAIIGSDAKIERCLMKDESCDPGYSDYVL 621
Qy 619 -EEG-----HSCVDL-----DD-----KGCFAEQ----- 646
Dy 622 QEEGPLQKSGKAVCRKCHPRCKTKTGYGFHQFQCECTGYKGEQCEDECPDYANEE 681
Qy 622 --ACQCPINCT-----HSCVDL-----DD-----KGCFAEQ----- 646
Dy 682 TRICLPCHQECRGCHGLDDHHCERNLKFEGDPYDNATTTVCVSNCPASHPYKRFPQEA 741
Qy 647 -----RASPLTSIVSAVVGILLVVLGVWVFGILI-----QVYKA 680
Dy 742 GKIGPYCSADSMQSGRLIEPTQVKIVMGVSMALILLCVFGIAFVLFSRHKNKDAVKM 801
Qy 681 NSKFIGITEL-PLTPSGAMPNQAQMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVK 739
Dy 802 TMLAGCEDSEPLRPSNVGNPLTKLRIKEAEIRRGVGLMGAFGRVFKGVWMPGESVK 861
Qy 740 IPVAIKVURENTSPANKIILDEAYVMAGVSPVSRLLGLCLTSTVOLVTLQMPYGLL 799
Dy 862 IPVAIKVLMEMSGSEKFELEEAYIMASVEHPNLLKLLAVCMTSQMWLITQLMPLGLCL 921
Qy 800 DHVRENRLGSOQLLNCMGIAGMSYLEVRLVHRDLAARNVLRKSPNHVKITDFGLA 859
Dy 922 DYVRNKKDKGSKALLNWSQIARGMAYLEERLVHRDLAARNVLVQTPSCVKITVGLA 981
Qy 860 RLLDIDETEXHADGKVPKIMWALESILRRRFTHOSDWSYGVTVWELMTGAKPYDGIP 919
Dy 982 KLLDFDSDYRAAGGNPKIKWALECIRHRVFTSKSDVWAFGITIWELLTYGARPYENV 1041
Qy 920 AREIPDLLEKGERLPQPPICITIDVYIMVIMKCMWIDSECRPFRELVEFSRMDPQRFV 979
Dy 1042 AKDPELIEIGHKLPQPDICSLDYVCILLSCWILDADARPTFKQLAETFAEKARDPGRYL 1101
Qy 980 VIONEDLPASPPLDSTFYRSLLEDDDDMGDLV----- 1010
Dy 1102 MI-----PGDKFMRLPSTYNQDEKDLITLAPVMAAAAAAAGASNDVPSTIA 1152
Qy 1011 DAEEYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEAAPS----- 1065
Dy 1153 ETDEYLOPKTRPSIMLPQPSA-----VEPS-DEMPKSLRYCK 1188
Qy 1066 -PLAP---SEGAGSDVFDGDLGMAAGKQLSLPTHDPSPLOQYSEDPTVPLPSETDGYVA 1121
Dy 1189 DPLAPDDETDGHEV-----GVGIR-----LNPLDDEDDYLM 1222
Qy 1122 PLTCSPOEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAGG 1181
Dy 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGVPA 1243
Qy 1182 AVENPEYL-----TPQGAAPQPPHPPAFSPAFDNLVYDQDPPERGAPSTFKGT 1232
Dy 1244 SVDNPEYLMGSTOAIAGLAQSMG--PHTPP-----PNTNGM 1280
Qy 1233 PTAENPE 1239
Dy 1281 PTHQHSQ 1287

RESULT 8
Q9UK79 ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
EMBL: AF177761; AAD56009.2; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; PEC1BE347E2D030C CRC64;

Query Match 27.6%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 3.4e-132;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Dy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYOGCVVQGNL 60
Qy 61 ELYLPTNASLFLQDIQEVQYVLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Dy 61 ELYLPTNASLFLQDIQEVQYVLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy 121 DPLNNTPTVGTGASPGGURELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Dy 121 DPLNNTPTVGTGASPGGURELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTVNRSPACHPCSPMKGSRGSGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQOC 240
Dy 181 LTLIDTVNRSPACHPCSPMKGSRGSGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQOC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGL 345
Dy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVTHSL 345

RESULT 9
Q8R2X1 ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```



```
Db 141 PEETATPKTGP--DHCCKAHPIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHP 197
Qy 629 NCHTSCVDLDDKCPAERASPLTSIVSAWV-GILLVVVLGVVFGILI--QYI---KANS 682
Db 198 NTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLCLVGVVGLIGLYLRRRHIVRKRTL 254
Qy 603 KFIGITEL--PLTPSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKI 740
Db 255 RLLORELVEPLTPSGEAPNOAHLRLKETEFKVKVLGSGAFGVYKGLWIPGEKVKI 314
Qy 741 PVAIKVLENTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLD 800
Db 315 PVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQMLPYGCLLD 374
Qy 801 HVRENKRGSLGDLNWCQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDPLGLAR 860
Db 375 YIREHKDNGISQYLLNWCQIAKGMVLEERLVRDLAARNVLKTPHVKITDPLGLAK 434
Qy 861 LLDIDETEHADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPA 920
Db 435 LLGADEKEYHABGGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPA 494
Qy 921 RIIPDLLEKGERLPOPPICITDVMIMVKWCMIDSECRPRFRELVSFSEMRARDPQFVV 980
Db 495 SEISSVLEKGERLPOPPICITDVMIMVKWCMIDADSRPKFRELIAEFSKWARDPPRYLV 554
Qy 981 IQ-NEDLGPASPLDTFYRSLLDDMDGLVDAAEYLVPOQGFPCPDAPAGAGGVHRRH 1039
Db 555 IOCDERMLPSTDSKFYRLMEEEDMEDIIVDAEYLVPHQGF-----598
Qy 1040 RSSSTRSGGDLTLGLEPSEEAAPRSL-----APSEGAGSDVFDGDLGMAAKQLSILP 1094
Db 599 NSPST-----SRTPLLSLSATSNSATNCID-----RNGQGH 632
Qy 1095 THDPSPLQRYSDPTVPLPSET--DGYVAPLTCSPQPEYVQNPQVRPQPPSPREGPLPA 1152
Db 633 VREDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS-----675
Qy 1153 RPAGATLERAKTSLSPKNGVVKDVF-----AFGGAIVENPEYL 1189
Db 676 -----TAMVQNIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL 715

RESULT 12
Q86714
ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhorak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SMO0219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
```

```
DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
SQ NON_TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 24.1%; Score 1635; DB 15; Length 567;
Best Local Similarity 53.7%; Pred. No. 2.9e-114;
Matches 345; Conservative 75; Mismatches 115; Indels 108; Gaps 17;

Qy 578 GPEADQCACAHYKDPFPCVAPRCSPGVKPDLSYMPKPFDEBEGACQPCPINCHTSCVDL 637
Db 1 GP-DHCCKAHPIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCRTCKGKP 57
Qy 638 DDKGCAPARQASPLTSIVSAWV-GILLVVVLGVVFGILI--QYI---KANSKFIGITEL- 690
Db 58 GLEGGP---NGSKTPSIAAGVVGGLCLVGVVGLIGLYLRRRHIVRKRTLRLRLRELV 114
Qy 691 -PLTPSGAMPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLE 749
Db 115 EPLTPSGEAPNOAHLRLKETEFKVKVLGSGAFGVYKGLWIPGEKVKIPVAIKELRE 174
Qy 750 NTPSKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENKRL 809
Db 175 ATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQMLPYGCLLDVIREHKNI 234
Qy 810 GSODLLNWCQIAKGSYLEDVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEY 869
Db 235 GSQYLLNWCQIAKGMVLEERLVRDLAARNVLKTPHVKITDPLGLAKLGADEKEY 294
Qy 870 HADGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 929
Db 295 HAEGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354
Qy 930 GERLPOPPICITDVMIMVKWCMIDSECRPRFRELVSFSEMRARDPQFVFIQ-NEDLGP 988
Db 355 GERLPOPPICITDVMIMVKWCMIDADSRPKFRELIAEFSKWARDPPRYLVIOGDERMHL 414
Qy 989 ASPLDSTFYRSLLDDMDGLVDAAEYLVPOQGFPCPDAPAGAGGVHRRHSSSTRSGG 1048
D. 415 PSPTDSKFYRLMEEEDMEDIIVDAEYLVPHQGF-----NSPST-----454
Qy 1049 GDLTLGLEPSEEAAPRSL-----APSEGAGSDVFDGDLGMAAKQLSPLTHDPSPLOR 1103
Db 455 -----SRTPLLSLSATSNSATNCID-----RNGQGHPRVREDSFVQR 492
Qy 1104 YSEDPTVPLPSET--DGYVAPLTCSPQPEYVQNPQVRPQPPSPREGPLPAARPAGATLER 1161
Db 493 YSSDPTGNFLEESIDGFL-----PAPEYVQ--LMPKKPS-----536
Qy 1162 AKTSLSPKNGVVKDVF-----AFGGAIVENPEYL 1189
Db 527 ---TAMVQNIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL 566

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag v-erb-A, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
```

1- SIMILARITY: BELONGS TO THE NUCLEAR (BY SIMILARITY).  
 DR ENBL; X52209; CAA36459.1; HSP; P10828; 2NLI.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000536; Hormone\_rec\_lig.  
 DR InterPro; IPR001723; Stdhmn\_receptor.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00105; zf-C4; 1.  
 DR PRINTS; PR00398; STRDHOMONER.  
 DR ProDom; PD00047; STROIDFINGER.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM000035; Znf\_pkinase; 1.  
 DR SMART; SM00430; HOLI\_C4steroid; 1.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00399; Znf\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_ATP; 1.  
 KW ATP-binding; DNA-binding; Nuclear\_Tyr; 1.  
 KW Transcription regulation; Nuclear protein; Receptor;  
 SQ SEQUENCE 962 AA; 108320 MW; 3CSAED791E4E95CE CRC64;  
 Query Match  
 Best Local Similarity 23.2%; Score 1570.5; DB 15; Length 962;  
 Matches 346; Conservative 75; Mismatches 145; Indels 125; Gaps 20;  
 QY 541 VEERVLQGLPRE-YVNAR-HCLP-----CHPEQ 568  
 DB 354 IEKQSYLLAEHYNYRKHNIHFWSKLLMKVADLRMIGAYHASRFLHMKVEPTLS 413  
 QY 569 PONGSVTCFGEADOCVACHYKDPFCVACPCGVKPDLSYMPIWKFDEGACQPCPI 628  
 DB 414 PQE-----VGP--DHCCKAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHP 465  
 QY 629 NCHTSCVDLDDKCPAEORASPLTSIVSAVV-GILLVVLGVVFGILI--QVI---KANS 682  
 DB 466 NCTCKGCGLEGCCP---NGSKTSPSIAAGVVGGLCLVVGILGLYLRHHVVKRTLR 522  
 QY 683 KFIGITEL--PLTPSGAMPNOAOMRILKETELRKVKVLGSGAGFYVYKGIWIPDGENVKI 740  
 DB 523 RLQERELVEPLTPSGEAPNOAHLRLKETEFKVKVLGFGAGFYVYKGLWIPGKVTI 582  
 QY 741 PVAIKVLRNTSPKANKEILDEAYVMAGVGSFYVSRLLGICLTSTVQLVLTOLMPYGCILLD 800  
 DB 583 PVAIKELREATSPKANKEILDEAYVMASVDPNHVCRLLGLCLTSTVQLITOLMPYGCILLD 860  
 QY 801 HVRENRLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLAR 860  
 DB 643 YIREHKDNIGSYLLNMCVQIAKGMNLYEERHVRDLAARNVVKTPQHVKITDFGLAK 702  
 QY 861 LLDIDETEHADGKVPKMALESILRRFTHOSDVWSYGVTVWELMTFGAKPYDGIPA 920  
 DB 703 QLCADKEVHAEGKVPKMALESILRRFTHOSDVWSYGVTVWELMTFGSKPYDGIPA 762  
 QY 921 REIPDLLEGERLPOPPICTIDVYIMVKCMWIDSECPREFELVSFSEARMARDPQRFV 980  
 DB 763 SEISSVLEKGERLPOPPICTIDVYIMVKCMWIDSECPREFELVSFSEARMARDPQRFV 980  
 QY 981 IQ-NEDLGAPSLDSTFYRSLLDDMGDLVDAEYLVYVQGFCCPDPAAGAGVHRRH 1039  
 DB 823 IQGDERMHLPSFTSKFRTLNEEDMEDIVDAEYLVPHQGF-----866  
 QY 1040 RSSSTRSCGGDLTLGLEPSEEEAPRSLAPSEGAGSDVFDGLGMGAAGKGLSLPTHDS 1099  
 DB 867 NSPST-----SRPLSSLSATN-----NSATKCIDRNGCH---898

QY 1100 PLQRYSEDPVLPSETDGYVAPLTCSPQEVYNQDVPQPSPREGPLPAARPA  
 DB 899 -----PVREDGFL-----PAPEVYNO--LMPKFPSTAMVQNYQIYNYI;  
 QY 1159 LERAKTLSPGKNGVYKVOVAFGGAIVENPEYL 1189  
 DB 938 ISKLPMDSRYN-----SHSTAVDNPEYL 961  
 RESULT 14  
 Q85468  
 ID Q85468 PRELIMINARY;  
 AC Q85468  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-MAR-2002 (TREMBLrel. 01, Last sequence update)  
 OS Avian erythroblastosis virus (T934) v-erbB gene.  
 OC Viruses; Retrovirus; Retroviridae; Avian type C retroviruses.  
 RX NCBI\_TaxID=11861;  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=88217326; PubMed=28971102;  
 RT "Common site of mutation in the erbB gene of avian erythroblastosis virus mutants that are temperature sensitive for transformation.";  
 RL Oncogene Res. 1:265-278(1987).  
 DR HSP; X06943; CAA30024.1;  
 DR HSP; P11362; IFGK  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_ATP; 1.  
 KW ATP-binding; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CCA0F8AF4 CRC64;  
 Query Match  
 Best Local Similarity 23.1%; Score 1562; DB 15; Length 545;  
 Matches 333; Conservative 72; Mismatches 124; Indels 98; Gaps 17;  
 QY 578 GPEADOCVACHYKDPFCVACPCGVKPDLSYMPIWKFDEGACQPCINCHTSCVDL 637  
 DB 1 GP--DHCCKAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHNCTRGCKP 57  
 QY 638 DDKCPAEORASPLTSIVSAVV-GILLVVLGVVFGILI--QVI---KANSKFIGITEL- 690  
 DB 58 GLEGCP---NGSKTSPSIAAGVVGGLCLVVGILGLYLRHHVVKRTLRRLQERELV 114  
 QY 691 -PLTPSGAMPNOAOMRILKETELRKVKVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLR 749  
 DB 115 EPLTPSGEAPNOAHLRLKETEFKVKVLGFGAGFYVYKGLWIPGKVTIPEKELRE 174  
 QY 750 NTSPPKANKEILDEAYVMAGVGSFYVSRLLGICLTSTVQLVLTOLMPYGCILLD 809  
 DB 175 ATSPKANKEILDEAYVMASVDPNHVCRLLGLCLTSTVQLITOLMPYGCILLD 809  
 QY 810 GSQDLNMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDET 869  
 DB 235 GSQYLLNMCVQIAKGMNLYEERHVRDLAARNVVKTPQHVKITDFGLAKOLGADEKEY 294  
 QY 870 HADGKVPKMALESILRRFTHOSDVWSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 929  
 DB 295 HABGKVPKMALESILRRFTHOSDVWSYGVTVWELMTFGSKPYDGIPASEISSVLEK 988  
 QY 930 GERLPOPPICTIDVYIMVKCMWIDSECPREFELVSFSEARMARDPQRFV 988  
 DB 355 GERLPOPPICTIDVYIMVKCMWIDSECPREFELVSFSEARMARDPQRFV 988



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.6984 Seconds  
(without alignments)  
4527.811 Million cell updates/sec

Title: SEQ4-653-675-12

Perfect score: 6776

Sequence: 1 MEALALCRWGLLLALLPPCA.....TFKGTPTAENPEVLGLDVPV 1247

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_101002.\*

1:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6608	97.5	1255	21	Human heregulin 2
2	6608	97.5	1255	22	Human tyrosine kin
3	6608	97.5	1255	22	HER2 transgene pla
4	6608	97.5	1255	23	Human HER2 (erbB2)
5	6602	97.4	1255	17	HER-2/neu protein.
6	6602	97.4	1255	20	Human HER-2/neu on
7	6602	97.4	1255	21	Human HER-2/neu pr
8	6602	97.4	1255	21	Amino acid sequenc
9	6602	97.4	1255	22	Human HER-2/neu pr
10	6602	97.4	1255	22	HER2/neu amino aci

11	6602	97.4	1255	23	AAE24067	Human Her-2 protei
12	6602	97.4	1255	23	AAE20479	Human Her-2/neu pr
13	6602	97.4	1255	23	AAW51143	Human Her-2/neu po
14	6602	97.4	1255	23	AAU77114	Human Her-2/neu po
15	6559	96.8	1433	14	AAK39568	Sequence of c-erbB
16	6438	95.0	1223	23	AAU98923	Human breast cance
17	6285	92.8	1200	21	AAE21208	Human HER-2/neu pr
18	5819.5	85.9	1256	21	AAE21199	Rat HER-2/neu prot
19	5819.5	85.9	1256	23	AAW51144	Rat Her-2/neu onco
20	5792.5	85.5	1256	21	AAE21206	Mouse Her-2/neu pr
21	5792.5	85.5	1256	22	AAE21206	Mouse Her-2/neu pr
22	5792.5	85.5	1256	23	AAW51151	Amino acid sequenc
23	4818	71.1	919	21	AAE21203	Mouse Her-2/neu on
24	4818	71.1	919	23	AAW51148	Human HER-2/neu fu
25	4068.5	60.0	920	23	AAW51152	Mouse Her-2/neu ex
26	4068.5	60.0	926	23	AAW51153	Mouse Her-2/neu ex
27	3702	54.6	712	21	AAE21204	Human HER-2/neu fu
28	3702	54.6	712	23	AAW51149	Her-2/neu extracel
29	3552	52.4	782	18	AAW19764	Her2-GM-CSF immuno
30	3550	52.4	653	21	AAE21200	Extracellular HER-
31	3550	52.4	653	23	AAW51145	Human Her-2/neu on
32	3512	51.8	645	22	AAE60408	Human ErbB2 oncopr
33	3512	51.8	645	22	AAE61593	Human ErbB2 extrac
34	3447	50.9	951	21	AAE44993	DC8cFV-erbB2EC fu
35	3344	49.4	624	11	AAE08222	Extracellular port
36	3066	45.2	1210	21	AAE19259	Amino acid sequenc
37	3066	45.2	1210	21	AAE50616	Human EGF receptor
38	3066	45.2	1210	23	AAE23019	Human Her-1 protei
39	3066	45.2	1210	23	AAW50768	Human epidermal gr
40	3064	45.2	1210	22	AAE68420	Amino acid sequenc
41	3053.5	45.1	654	21	AAE21205	Rat HER-2/neu prot
42	3053.5	45.1	654	23	AAW51150	Rat Her-2/neu onco
43	3025	44.6	1210	23	ABP51768	Human epidermal gr
44	2974	43.9	583	23	AAE20483	Human protein for
45	2974	43.9	587	23	AAE20481	Human protein for

#### ALIGNMENTS

RESULT 1  
AA92620  
ID AA92620 standard; Protein; 1255 AA.

XX AA92620;

XX 10-AUG-2000 (first entry)

XX Human heregulin 2 (Her2).

DE Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
KW self-protein; cancer; breast cancer; prostate cancer;  
KW cell-associated peptide antigen; foreign epitope.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..173

FT /label= "N-terminal

FT /note= "mature polypeptide"

FT Region 5..25

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Region 59..73

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Region 103..117

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Region 149..163

FT /label= insertion region

FT /note= "suitable for foreign epitope insertion"

FT Domain 174..323

FT Region /label= Cysteine\_rich\_domain  
210..224  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
250..264  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
324..433  
FT /label= Ligand\_binding\_domain  
325..339  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
369..383  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
465..479  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
484..623  
FT /label= Cysteine\_rich\_domain  
579..593  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
624..654  
FT /label= Transmembrane\_domain  
632..652  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
653..667  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
655..1010  
FT /label= Tyrosine\_kinase\_domain  
661..675  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
695..709  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
710..730  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
1011..1235  
FT /label= C-terminal\_domain  
XX  
PN WO200020027-A2.  
XX  
PD 13-APR-2000.  
XX  
PF 05-OCT-1999; 99WO-DK00525.  
XX  
PR 05-OCT-1998; 98DK-0001261.  
PR 20-OCT-1998; 98US-0105011.  
XX  
PA (MEBI-) M & E BIOTECH AS.  
XX  
PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
PI Gautam A, Birk P, Karlsson G;  
XX  
DR WPI: 2000-349917/30.  
DR N-PSDB; AAA09455.  
XX  
XX Inducing immune responses to weakly immunogenic, tumor associated  
PT peptide antigens for the treatment of breast and prostate cancer  
XX  
PS Claim 62; Page 193-198; 220pp; English.  
XX  
CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
CC response. Subdominant CTL epitopes, antibody binding regions and  
CC cysteine residues involved in disulfide bonds are preserved in the  
CC immunogenized forms. Regions suitable for the insertion of foreign T  
CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
CC cell-associated peptide antigens (PA) such as those associated with  
CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
CC The method comprises effecting simultaneous presentation by antigen  
CC producing cells (APCs) of the animals immune system of: (1) at least 1  
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
CC B-cell group derived from the cell-associated PA; and (2) at least 1  
CC first T helper cell group which is foreign to the animal. Analogues of  
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
CC part of all known and predicted CTL and B-cell epitopes of the respective  
CC PA and including at least one foreign T helper epitope are also claimed.  
CC The method is used to treat prostate, prostate/breast or breast cancer  
CC when the PA is human PSM, FGF8b and Her2, respectively.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6608; DB 21; Length 1255;  
Best Local Similarity 97.4%; Pred. No. 0;  
Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;  
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHDMLRHLHYQCQVVGQNL 60  
Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKLRLPASPETHDMLRHLHYQCQVVGQNL 60  
Qy 61 ELTYLPTNASLSFLQDIQEVGYVLI AHNOVRQVPLQRLIRVGTQLPEDNYALVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVGYVLI AHNOVRQVPLQRLIRVGTQLPEDNYALVLDNG 120  
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQIRNPOLCYQDTILWKDIFHKNNQLA 180  
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQIRNPOLCYQDTILWKDIFHKNNQLA 180  
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVAGGCARCKGPIPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVAGGCARCKGPIPTDCCHQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREAVRATSAN 360  
Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLEITGYLISAWPDSLP 420  
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFTLEITGYLISAWPDSLP 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLALIHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLALIHNNTHLCFVHTV 480  
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCNCVSQFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCNCVSQFLRGQEC 540  
Qy 541 VECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
Db 541 VECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
Qy 601 PSGVKPDLSPYMPITWKPDEEGACQPCINCTHSCVDLDDKGCPAEQASPLTSIVSVAVG 660  
Db 601 PSGVKPDLSPYMPITWKPDEEGACQPCINCTHSCVDLDDKGCPAEQASPLTSIVSVAVG 660  
Qy 661 ILLVWLGVVFGILI-----QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712  
Db 661 ILLVWLGVVFGILIKERQKKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Qy 713 RKVKVLGSGAFGTGKIWIIPDGENVKIPVAIKVLRENTSPKANKELIDENYVAGVGP 772  
Db 721 RKVKVLGSGAFGTGKIWIIPDGENVKIPVAIKVLRENTSPKANKELIDENYVAGVGP 780



```
QY 773 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRGRGLSGDILLNWCQIAKGMHSYLEDVR 832
DB 781 YVSRLLGICLTSTVOLVTLQMPYGCCLLDHVRENRGRGLSGDILLNWCQIAKGMHSYLEDVR 840
QY 833 LVHRDLAARNVLVKSPNHVKITDFGLARLLIDIDETEHADGGKVPKMMALESILRRRFT 892
DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLIDIDETEHADGGKVPKMMALESILRRRFT 900
QY 893 HQSDVMSYGVTVWELMTFGAKPYDGI:PAEIPOLLEKGERLPPOPICTIDVYIMVKCWM 952
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI:PAEIPOLLEKGERLPPOPICTIDVYIMVKCWM 960
QY 953 ISECEPRRELVSERWARDPQRFVWQNEIDLGPASPLDSTFYRSLLEDDDMGLVDA 1012
DB 961 ISECEPRRELVSERWARDPQRFVWQNEIDLGPASPLDSTFYRSLLEDDDMGLVDA 1020
QY 1013 EEVLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1072
DB 1021 EEVLVPQOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSEG 1080
QY 1073 AGSDVFDGLGMAAGLQSLPHTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1132
DB 1081 AGSDVFDGLGMAAGLQSLPHTHDPSPLOQYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140
QY 1133 NOPDVRPQPSREGPLPAARPAAGATLERAKTLSPCKNGVVKDVFAGGAVENPEYLTPO 1192
DB 1141 NOPDVRPQPSREGPLPAARPAAGATLERAKTLSPCKNGVVKDVFAGGAVENPEYLTPO 1200
QY 1193 GGAAPOPHPPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1247
DB 1201 GGAAPOPHPPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1255
```

## RESULT 2

```
AAE12130
ID AAE12130 standard; Protein; 1255 AA.
XX
AC AAE12130;
XX
DT 18-DEC-2001 (first entry)
XX
DE Human tyrosine kinase-type receptor, HER-2.
XX
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
XX
OS Homo sapiens.
```

```
XX Key Location/Qualifiers
FH Region 774..782
FT /note= "Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX
XX 16-MAR-2000; 2000US-0527487.
XX
XX (GENZ ) GENZYME CORP.
XX
XX Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX
XX N-PSDB; AAD19731.
XX
```

Novel synthetic therapeutic compound for inducing immune response and for use in adoptive immunotherapy, has enhanced binding to major histocompatibility molecules and enhanced immunoregulatory properties

```
XX
PS
XX
CC The invention relates to synthetic therapeutic compounds (antigenic
CC peptides) with enhanced binding to major histocompatibility complex
CC (MHC) molecules and enhanced immunoregulatory properties relative
CC to their natural counterparts. Compounds of the invention are useful
CC for inducing an immune response in a subject and for use in adoptive
CC immunotherapy. They are useful as components of anti-cancer vaccines
CC and to expand immune effector cells that are specific for cancers
CC characterised by expression of the breast cancer antigen, HER-2.
CC Polynucleotides that encode peptides of the invention are useful as
CC hybridisation probes and as primers for the detection of genes of gene
CC transcripts that are expressed in antigen presenting cells (APCs), to
CC confirm transduction of polynucleotides into host cells. The present
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds
CC of the invention are designed based on the HER-2 antigenic peptide
CC (774-782).
XX
SQ Sequence 1255 AA;
```

```
Query Match 97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;

QY 1 MELAALCRMGLLLALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVQGCVOVQGNL 60
DB 1 MELAALCRMGLLLALLPPGAASQVCTGTDMLRLPASPEHDLMLRHLVQGCVOVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIHNOVROVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHKNNOLA 180
QY 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC 240
DB 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLALHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNHQVTAEDGTORCEKSKPCARVCYGLGMOYKANSKFTGIT 360
DB 301 YNYLSTDVGSCTLVCPHLNHQVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420
DB 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLSLRELGLSGSLAIHHNTHLCFVHT 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLSLRELGLSGSLAIHHNTHLCFVHT 480
QY 481 PWDQLFRNPHQALLHTANRPEDECEGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
DB 481 PWDQLFRNPHQALLHTANRPEDECEGLACHQLCARGHCWGPGPTQCVNCSQFLRQEC 540
QY 541 VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHKDPFCVCARC 600
DB 541 VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHKDPFCVCARC 600
QY 601 PSGVKPDLISYMPINWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
DB 601 PSGVKPDLISYMPINWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSIVSAVVG 660
QY 661 ILLVVVLGVVFGILII----QYIKANS--KFIGITEL--LPTSGAMPNQAQMRILKETEL 712
DB 661 ILLVVVLGVVFGILIIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720
```

```
QY 713 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 772
DB 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 773 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEVDV 832
DB 781 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEVDV 840
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 892
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 900
QY 893 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 952
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960
QY 953 IDSECRPRFRELSEFRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDDDMGDLVDA 1012
DB 961 IDSECRPRFRELSEFRMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDDDDMGDLVDA 1020
QY 1013 EYLVPOOGFFCDDPAPGAGMVHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072
DB 1021 EYLVPOOGFFCDDPAPGAGMVHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080
QY 1073 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132
DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1133 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1192
DB 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTQP 1200
QY 1193 GGAAPQHPAPSPAFDNLVYDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1247
DB 1201 GGAAPQHPAPSPAFDNLVYDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN W0200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX
PA (GETH ) GENENTECH INC.
XX
PI Erickson S, Schwall R;
XX
DR WPI; 2001-061962/07.
XX
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
```

```
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 97.5%; Score 6608; DB 22; Length 1255;
Best Local Similarity 97.4%; Pred. No. 0;
Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;
QY 1 MELAALCRWGLLALLPPGAASTOVCTGTDMKRLPASPETHLMLRHLHYGCVQVQGNL 60
DB 1 MELAALCRWGLLALLPPGAASTOVCTGTDMKRLPASPETHLMLRHLHYGCVQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLQDIQEVGYVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNQOLA 180
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNQOLA 180
QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
DB 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCITLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLHMOYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCITLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLHMOYIKANSKFIGIT 360
QY 361 ELEFAGCKITFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLSEITGILYLSAWPDSL 420
DB 361 IQEFAGCKITFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLSEITGILYLSAWPDSL 420
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLOGIGISWGLRSRLRELGLALIHNNTHLCFVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLOGIGISWGLRSRLRELGLALIHNNTHLCFVHTV 480
QY 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTOCVNCSCQFLRGQEC 540
DB 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTOCVNCSCQFLRGQEC 540
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600
QY 601 PSGVKPDLSTYMPIWKPPDEEGACQPCINCHTSCVDLDDKGCPEAQSPUTSIVSAVVG 660
DB 601 PSGVKPDLSTYMPIWKPPDEEGACQPCINCHTSCVDLDDKGCPEAQSPUTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIIKRRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 712
DB 661 ILLVVLGVVFGILIIKRRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 713 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 772
DB 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780
QY 773 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEVDV 832
DB 781 YVSRLLGICLTSTVOLVQLMPYGCCLLDHVRENRGRIGSDLLNWCQIAKMSYLEVDV 840
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 892
DB 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPICKMALESILRRFT 892
```

Db 841 LVHRDLAARNVVKSPNHVXITDFGLARLLDIDETEHADGGKVPKMKMALESILRRFT 900  
 QY 893 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLEKGERLPQPPICITIDVYIMVKWM 952  
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLEKGERLPQPPICITIDVYIMVKWM 960  
 QY 953 IDSECRPRFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGDLVDA 1012  
 Db 961 IDSECRPRFELVSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGDLVDA 1020  
 QY 1013 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1072  
 Db 1021 EYLVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080  
 QY 1073 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1132  
 Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYV 1140  
 QY 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1192  
 Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200  
 QY 1193 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTTAENPEYLGDLVPV 1247  
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTTAENPEYLGDLVPV 1255

## RESULT 4

AAU74545

ID AAU74545 standard; Protein; 1255 AA.

XX AAU74545;

AC AAU74545;

XX AAU74545;

DT 23-APR-2002 (first entry)

DE Human HER2 (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;

KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;

KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;

KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;

KW glial disorder; astrocytal disorder; hypothalamic disorder;

KW glandular disorder; macrophagal disorder; epithelial disorder;

KW stromal disorder; blastocoealic disorder; inflammatory disorder;

KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

OS Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth

XX factor receptor, ErbB or cancer in mammal, comprises administering

XX anti-ErbB antibody-maytansinoid conjugate to the mammal

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

XX

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.  
 xx

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6608; DB 23; Length 1255;

Best Local Similarity 97.4%; Pred. No. 0;

Matches 1222; Conservative 8; Mismatches 17; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLALLPPGAASQTQCTGTDMLRLPASPTHLDMLRHLVGGCVQVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQTQCTGTDMLRLPASPTHLDMLRHLVGGCVQVQGNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQGVYLAHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILWKDIFHKNQLA 180

Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRLNPOLCYQDTILWKDIFHKNQLA 180

QY 181 LTLIDNTRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240

Db 181 LTLIDNTRSRACHPCSPMKSGRCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMYIKANSKFIT 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKCKPCARVCYGLGMYIKANSKFIT 360

QY 361 ELEFAGCKIFGSLAFLEPSEFDGDPASNTAPLOEQVFEETLEETIGYLIYISAWPDSL 420

Db 361 IQEFAGCKIFGSLAFLEPSEFDGDPASNTAPLOEQVFEETLEETIGYLIYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480

QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVNCSQFLRQEC 540

QY 541 VEECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVCAHYKDPFCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFGEADQCVCAHYKDPFCVARC 600

QY 601 PSGVKPDLSTYMPIWKFPPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVVG 660

Db 601 PSGVKPDLSTYMPIWKFPPDEEGACQPCINCTHSCVDLDDKGCPEORASPLTSIVSAVVG 660

QY 661 ILLVVVLGVVFGILI----QYIKANS--KFIGITEL--PLTSGAMPNQAOMRILKETEL 712

Db 661 ILLVVVLGVVFGILI----QYIKANS--KFIGITEL--PLTSGAMPNQAOMRILKETEL 712

QY 713 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYVMAGVGS 772

Db 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVRENTSPKANKIILDEAYVMAGVGS 780

QY 773 YVSRLLGICLTSTVQLVTQLMPYGLLDHVRNRRGRGLSQDILLNMCWQIAGKMSYLEYR 832

Db 781 YVSRLLGICLTSTVQLVTQMPYGCLLDHRVRENRGLQSDDLWNCQIAKGMSTYLEDVR 840  
QY LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 892  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWKMALESILRRFT 900  
QY HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCW 952  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCW 960  
QY IDSECRPRFRELVSSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1012  
Db 961 IDSECRPRFRELVSSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020  
QY 1013 EYLVPQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072  
Db 1021 EYLVPQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
QY 1073 AGSDVFDGDLGMAAKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132  
Db 1081 AGSDVFDGDLGMAAKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1133 NOPDVRQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFATGGAVENPEYLTPO 1192  
Db 1141 NOPDVRQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFATGGAVENPEYLTPO 1200  
QY 1193 GGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVPV 1247  
Db 1201 GGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVVPV 1255

## RESULT 5

AAW01111  
ID AAW01111 standard; Protein; 1255 AA.  
XX AAW01111;  
AC AAW01111;  
DT 01-JAN-1997 (first entry)  
XX HER-2/neu protein.  
DE HER-2/neu protein.  
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; immunisation; tumour; vaccine; vector.  
XX Homo sapiens.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Domain 676..1255  
FT /label= Intracellular domain  
FT /note= "claimed domain, useful for immunisation"  
XX W09630514-A1.  
XX 03-OCT-1996.  
XX 28-MAR-1996; 96WO-US01689.  
XX 31-MAR-1995; 95US-0414417.  
XX (UNIW ) UNIV WASHINGTON.  
XX Cheever MA, Disis ML;  
XX WPI; 1996-455361/45.  
XX N-PSDB; AAT40739.  
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or  
PT treatment of malignancies with which the HER-2/neu oncogene is  
PT associated  
XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.

XX SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 17; Length 1255;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLKRLPASPTHLMRLHYGCQVVGQNL 60  
Db 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLKRLPASPTHLMRLHYGCQVVGQNL 60  
QY 61 ELTYLPTNASISFLQDIOEVQGYVLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLNG 120  
Db 61 ELTYLPTNASISFLQDIOEVQGYVLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLNG 120  
QY 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180  
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGGCARCKGPIPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVCAAGGCARCKGPIPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYRTTGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGYRTTGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVTASAN 360  
QY 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLEITGTYLISAWPDSLP 420  
Db 361 IQEPAGCKKIFGSLAFIPESPDGPASNTAPLQPEQLQVFTLEITGTYLISAWPDSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGSLWGLRSILRELGSGLALHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGSLWGLRSILRELGSGLALHNNTHLCFVHTV 480  
QY 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVARC 600  
QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660  
Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660  
QY 661 ILLVVVLGVVFGILI-----QYIKANS--KFTIGITEL--PLTPSGAMPNQAQMRILKETEL 712  
Db 661 ILLVVVLGVVFGILI KRRQKKIRKYTMRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 713 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVWAGVSP 772  
Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVWAGVSP 780  
QY 773 YVSRLLGICLTSTVQLVTQMPYGCLLDHRVRENRGLQSDDLWNCQIAKGMSTYLEDVR 832  
Db 781 YVSRLLGICLTSTVQLVTQMPYGCLLDHRVRENRGLQSDDLWNCQIAKGMSTYLEDVR 840

QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 892  
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900  
 QY 893 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 952  
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960  
 QY 953 IDSECRPRFRELVSERMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1012  
 DB 961 IDSECRPRFRELVSERMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1020  
 QY 1013 EYLVPQGGFCFDPAPACAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1072  
 DB 1021 EYLVPQGGFCFDPAPACAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
 QY 1073 AGSDVFDGDLGMAAGLQSLTHDPSPQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132  
 DB 1081 AGSDVFDGDLGMAAGLQSLTHDPSPQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
 QY 1133 NQDVRPQPPSPREGPLPAARPAGATLERAKTLSFGKNGVVKOVFAFGGAVENPEYLTPO 1192  
 DB 1141 NQDVRPQPPSPREGPLPAARPAGATLERAKTLSFGKNGVVKOVFAFGGAVENPEYLTPO 1200  
 QY 1193 GGAAPQHPAPAFSPAFDNLYYWDODPPERCAPPSTFKGTPTAENPEYLGLDVVP 1247  
 DB 1201 GGAAPQHPAPAFSPAFDNLYYWDODPPERCAPPSTFKGTPTAENPEYLGLDVVP 1255

## RESULT 6

AAW92406

ID AAW92406 standard; Protein; 1255 AA.

XX AAW92406;

XX 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;

XX malignancy; treatment; tumour.

XX Homo sapiens.

XX Key Location/Qualifiers

FH Region 676..1255

FT /note= "region which elicits immune response"

XX US5869445-A.

XX 09-FEB-1999.

XX 01-APR-1996; 96US-0625101.

XX 01-APR-1996; 96US-0625101.

XX 17-MAR-1993; 93US-0033644.

XX 12-AUG-1993; 93US-0106112.

XX 31-MAR-1995; 95US-041417.

XX (UNIW ) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1999-152835/13.

XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to

XX an HER-2/neu associated malignancy, particularly for treating or

XX preventing tumours

XX Claim 3; Column 31-38; 26pp; English.

XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC

CC of this protein is used in a method for eliciting or enhancing an immune  
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
 CC B cells to produce an immune response to the HER-2/neu protein. The  
 CC method can be used for immunisation against a malignancy in which the  
 CC HER-2/neu oncogene is associated and in the treatment of an existing  
 CC tumour, or to prevent tumour occurrence or reoccurrence.

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 20; Length 1255;  
 Best Local Similarity 97.2%; Pred. No. 0;  
 Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;  
 QY 1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMLRHLHYGQCQVVGNL 60  
 DB 1 MELAALCRWGLLLALLPFGAASCTGCTDMKRLPASPETHLDMLRHLHYGQCQVVGNL 60  
 QY 61 ELTYLPTNASLSFLQDIEQVGYVLI AHNQVRQVPLORLRIVRGTQLPEDNYALAVLNG 120  
 DB 61 ELTYLPTNASLSFLQDIEQVGYVLI AHNQVRQVPLORLRIVRGTQLPEDNYALAVLNG 120  
 QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180  
 DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180  
 QY 181 LTLIDTNRSRACHPCSPMKGSRGWCSESSDCQSLTRTVACAGCARCKGPLTDCCHQC 240  
 DB 181 LTLIDTNRSRACHPCSPMKGSRGWCSESSDCQSLTRTVACAGCARCKGPLTDCCHQC 240  
 QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300  
 DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300  
 QY 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGHWYIKANSKFGIT 360  
 DB 301 YNYLSTDVSGCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLEVRVAVTSAN 360  
 QY 361 ELEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETLGYLVIASWPDSL 420  
 DB 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETLGYLVIASWPDSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLIGISWGLRSLRELGLGLAIHHNTHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLIGISWGLRSLRELGLGLAIHHNTHLCFVHTV 480  
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQIARGHCWGPGPTQCVNCSQFLRGQEC 540  
 DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGACHQIARGHCWGPGPTQCVNCSQFLRGQEC 540  
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQVCAHAHYKDPFCVAVC 600  
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQVCAHAHYKDPFCVAVC 600  
 QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660  
 DB 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVG 660  
 QY 661 ILLVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712  
 DB 661 ILLVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 720  
 QY 713 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEITLDEAYMAGVSP 772  
 DB 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEITLDEAYMAGVSP 780  
 QY 773 YVSRLLGICLSTVQLVTLQMPYGCILLDHVRENRGRIGSQDLLNWCQIAKGSYLEDVR 832  
 DB 781 YVSRLLGICLSTVQLVTLQMPYGCILLDHVRENRGRIGSQDLLNWCQIAKGSYLEDVR 840  
 QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 892  
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 952  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 960  
Qy 953 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1012  
Db 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1020  
Qy 1013 EYLVPQGGFCPPDPAFGAGMWHRRHRSSTSRGGDLTLGLSPSEEEARSPPLASEG 1072  
Db 1021 EYLVPQGGFCPPDPAFGAGMWHRRHRSSTSRGGDLTLGLSPSEEEARSPPLASEG 1080  
Qy 1073 AGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1132  
Db 1081 AGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGVYVAPLTCSPQPEYV 1140  
Qy 1133 NQPDVRFPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGAVENPEYLTTPQ 1192  
Db 1141 NQPDVRFPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVFAFGAVENPEYLTTPQ 1200  
Qy 1193 GGAAPQHPHPPAFSPADNLYWDDPPRGAPSTFKGTPTAENPEYLGIDVPV 1247  
Db 1201 GGAAPQHPHPPAFSPADNLYWDDPPRGAPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 7

AA821198  
ID AAB21198 standard; protein; 1255 AA.  
XX AAB21198;  
AC AAB21198;  
DT 12-JAN-2001 (first entry)  
XX Human HER-2/neu protein.  
DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX Homo sapiens.  
XX WO200044899-A1.  
XX 03-AUG-2000.  
XX 28-JAN-2000; 2000WO-US02164.  
XX 29-JAN-1999; 99US-0117976.  
XX (CORI-) CORIXA CORP.  
XX (SMIK ) SMITHKLINE BEECHAM.  
XX Cheever MA, Gheysen D;  
XX WPI; 2000-505976/45.  
XX N-PSDB; AAA89736.  
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
XX Sequence 1255 AA;  
SQ Query Match  
Best Local Similarity 97.4%; Score 6602; DB 21; Length 1255;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;  
Qy 1 MELAALCRWGLLLALLPFGAASTOVCTGTDMLKRLPASPETHLDMRLHYGQCQVQGNL 60  
Db 1 MELAALCRWGLLLALLPFGAASTOVCTGTDMLKRLPASPETHLDMRLHYGQCQVQGNL 60  
Qy 61 ELTYLPTNASLFLQDIQEVGYVLIAHNOVQVPLQRLRIVRGTLQEDFNALAVLNG 120  
Db 61 ELTYLPTNASLFLQDIQEVGYVLIAHNOVQVPLQRLRIVRGTLQEDFNALAVLNG 120  
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTELKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTELKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVAGGCARCKGPLPTDCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGWGESSEDCQSLTRTVAGGCARCKGPLPTDCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360  
Qy 361 ELEFAGCKIFGSLAFIPESPFGDPASNTAPLQPEQLQVFTLEBETCYLYISAMPDSL 420  
Db 361 IQEFAGCKIFGSLAFIPESPFGDPASNTAPLQPEQLQVFTLEBETCYLYISAMPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLGSLGLALHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLGSLGLALHNNHLCFVHTV 480  
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSOFURGQEC 540  
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSOFURGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPECVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPPECVARC 600  
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCETHSCVDLDDKCPAEORASPLTSIISAVVG 660  
Db 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCETHSCVDLDDKCPAEORASPLTSIISAVVG 660  
Qy 661 ILLVVVLGVVFGILI---QYIKANS--KFTIGITEL--PLTPSGAMPNQAOIRILKETEL 712  
Db 661 ILLVVVLGVVFGILI---QYIKANS--KFTIGITEL--PLTPSGAMPNQAOIRILKETEL 712  
Qy 713 RKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGP 772  
Db 721 RKVKVLSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVGP 780  
Qy 773 YVSRLLGLCLTSTVQLVTQMLPYCCLLDHVRENRGRIGSQDLNWKCMQIAKMSYLEVDV 832  
Db 781 YVSRLLGLCLTSTVQLVTQMLPYCCLLDHVRENRGRIGSQDLNWKCMQIAKMSYLEVDV 840  
Qy 833 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHYHADGKVPDKWMALESILURRFT 892  
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHYHADGKVPDKWMALESILURRFT 900  
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 952  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCM 960  
Qy 953 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1012

Db 961 IDSECRPRELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020  
QY 1013 EYLVLPQOQFFCDDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEEAPRSLAPSEG 1072  
Db 1021 EYLVLPQOQFFCDDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEEAPRSLAPSEG 1080  
QY 1073 AGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1132  
Db 1081 AGSDVFDGLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1133 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTQ 1192  
Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTQ 1200  
QY 1193 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1247  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 8  
AAY84780  
ID AAY84780 standard; Protein; 1255 AA.  
XX AAY84780;  
AC AAY84780;  
DT 08-AUG-2000 (first entry)  
XX Amino acid sequence of the SPLICE erbB-2 receptor protein.  
DE SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KW bone resorption; inflammatory disease; degenerative disorder;  
KW wound healing.  
XX Homo sapiens.  
XX  
XX WO200020579-A1.  
XX  
XX 13-APR-2000.  
XX  
XX 01-OCT-1999; 99WO-CA00912.  
XX  
XX 02-OCT-1998; 98US-0165192.  
XX  
XX (UYMC-) UNIV MCMASTER.  
XX  
XX Muller WJ, Siegel PM;  
XX  
XX WPI; 2000-303768/26.  
DR N-PSDB; AAA14812.  
XX  
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE  
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -  
XX  
PS Claim 3; Fig 2; 60pp; English.  
CC The present sequence represents a SPLICE erbB-2 receptor protein. The  
CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
CC conserved cysteine residues, compared to the unspliced protein. The  
CC erbB-2 polynucleotide is used to construct probes for detecting  
CC disorders of cell transformation such as cancer. Antibodies to the  
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents  
CC (e.g. antisense oligonucleotides) which inhibit the expression of  
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and  
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful  
CC for treating conditions involving damaged cells including conditions  
CC in which degeneration of tissue occurs, such as arthropathy, bone  
CC resorption, inflammatory diseases, degenerative disorders of the  
CC central nervous system and wound healing.  
XX  
XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 21; Length 1255;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;  
QY 1 MELAALCRWGLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRHLVQGCQVOQGNL 60  
Db 1 MELAALCRWGLLALLPPGAASTQVCTGTDMLKRLPASPTHLDMLRHLVQGCQVOQGNL 60  
QY 61 ELTYLPTNASLFLQDIQEVQGVVLIAHNOVROVPLQRLIRVGTOLFEONYALAVLDNG 120  
Db 61 ELTYLPTNASLFLQDIQEVQGVVLIAHNOVROVPLQRLIRVGTOLFEONYALAVLDNG 120  
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180  
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNQLA 180  
QY 181 LTLIDNTRGRACHPCSPMCKGSRGWESSEDCOSLTRTVCCAGGCARCKGPLPDDCCHQEC 240  
Db 181 LTLIDNTRGRACHPCSPMCKGSRGWESSEDCOSLTRTVCCAGGCARCKGPLPDDCCHQEC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKCSKPCARVCYGLGMYIKANSKFITG 360  
Db 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQCEKCSKPCARVCYGLGMYIKANSKFITG 360  
QY 361 ELEFAGCKIKFGLAFDPESFGDGPASNTAPLOPEQLQVPEETLEETGYIYISAWPDSLP 420  
Db 361 IQEFAGCKIKFGLAFDPESFGDGPASNTAPLOPEQLQVPEETLEETGYIYISAWPDSLP 420  
QY 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLFCFVHTV 480  
Db 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNTHLFCFVHTV 480  
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVNCSOFLRQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVNCSOFLRQEC 540  
QY 541 VEECRVQLGLPREYVNHARCLPCHPECQONGSVTCFGEADQCVCAHYKDPFPCVARC 600  
Db 541 VEECRVQLGLPREYVNHARCLPCHPECQONGSVTCFGEADQCVCAHYKDPFPCVARC 600  
QY 601 PSGVKPDLSYMPIWKFPPDEEGACQPCINCTHSCVDLDDKGCPEABORASPLTSIVSAVG 660  
Db 601 PSGVKPDLSYMPIWKFPPDEEGACQPCINCTHSCVDLDDKGCPEABORASPLTSIVSAVG 660  
QY 661 ILLVVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNOAMRILKETEL 712  
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNOAMRILKETEL 720  
QY 713 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKSILDEAYVMAGVSP 772  
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKSILDEAYVMAGVSP 780  
QY 773 YVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRLGSGQDLLNWCQIAGKSYLEDVR 832  
Db 781 YVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRLGSGQDLLNWCQIAGKSYLEDVR 840  
QY 833 LVHRLDAARNVLKSPNKHVITDFGLARLLDIDETEHADGGKVPKKNWALESIKRRFT 892  
Db 841 LVHRLDAARNVLKSPNKHVITDFGLARLLDIDETEHADGGKVPKKNWALESIKRRFT 900  
QY 893 HQSDVMSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDVMVMKCMW 952  
Db 901 HQSDVMSVGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDVMVMKCMW 960  
QY 953 IDSECRPRELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1012  
Db 961 IDSECRPRELVSEFSRWARDPQRFVWIONEDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020  
QY 1013 EYLVLPQOQFFCDDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEEAPRSLAPSEG 1072



```
Db 1021 EEVLVPOQGFCCDPAPGAGGVHRRSSSTRSGGDLTLGLEPEEEAPRPLAFSEG 1080
QY 1073 AGSDVFDGDLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGDLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1140
QY 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1192
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1200
QY 1193 GGAAPQHPPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1247
Db 1201 GGAAPQHPPPAFSPAFDNLYYDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN W0200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX
PI Cheever MA, Hand-Zimmermann S;
XX
PI WPI: 2001-476112/51.
XX
DR N-PSDB; AAH23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer -
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 22; Length 1255;
Best Local Similarity 97.2%; Pred. No. 0;
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MEALACRWGLLALLPPGAASQVCTGTDWKLRPASPETHLDMLRHLVQGCQVQGNL 60
Db 1 MEALACRWGLLALLPPGAASQVCTGTDWKLRPASPETHLDMLRHLVQGCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVGVVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
```

```
Db 61 ELTYLPTNASLSFLQDIQEVGVVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNOLA 180
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNOLA 180
QY 181 LTLIDNRSRACHPCSPMKCKSGCWGSESDCSLRTTVCCAGGCARCKGPLTCCCHQEC 240
Db 181 LTLIDNRSRACHPCSPMKCKSGCWGSESDCSLRTTVCCAGGCARCKGPLTCCCHQEC 240
QY 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKGSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRRAVTSAN 360
Db 301 YNYLSTDVGSCTLVCPLHNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRRAVTSAN 360
QY 361 ELEPAGCKITFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITCYLVISAWPDSL 420
Db 361 IQEPAGCKITFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITCYLVISAWPDSL 420
QY 421 DLSVFQNLQVIRGRIHLHGAYSLTLQGLGISWLGRLSLRELGLALIHNTHLFCFVHTV 480
Db 421 DLSVFQNLQVIRGRIHLHGAYSLTLQGLGISWLGRLSLRELGLALIHNTHLFCFVHTV 480
QY 481 PWDLFRNPHQALLHTANRPEDEBCVGEGLACHQLCARGHCWGPGPTQCVCNSQFLRGQEC 540
Db 481 PWDLFRNPHQALLHTANRPEDEBCVGEGLACHQLCARGHCWGPGPTQCVCNSQFLRGQEC 540
QY 541 VEESRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
Db 541 VEESRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600
QY 601 PSGVKPDLSTYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
Db 601 PSGVKPDLSTYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660
QY 661 ILLVVLGVVFGILJ----QYIKANS--KFIGITEL--PLTTPGAMPNQAMRILKETEL 712
Db 661 ILLVVLGVVFGILIKRRQKIRKYTMRLLQETELVEPLTPSGAMPNQAMRILKETEL 720
QY 713 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKETILDEAYVMAGVCS 772
Db 721 RKVKVLSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSKANKETILDEAYVMAGVCS 780
QY 773 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSDLLNWCMIKAGMSYLEDVR 832
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGRIGSDLLNWCMIKAGMSYLEDVR 840
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRRFT 892
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRRFT 900
QY 893 HQSDVMSYGVTVWELMTFFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYIMVVKCM 952
Db 901 HQSDVMSYGVTVWELMTFFGAKPYDGPAREIPDLLEKGERLPQPICTIDVYIMVVKCM 960
QY 953 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGLVDA 1012
Db 961 IDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDGLVDA 1020
QY 1013 EYLVPOQGFCCDPAPGAGGVHRRSSSTRSGGDLTLGLEPEEEAPRPLAFSEG 1072
Db 1021 EYLVPOQGFCCDPAPGAGGVHRRSSSTRSGGDLTLGLEPEEEAPRPLAFSEG 1080
QY 1073 AGSDVFDGDLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1132
Db 1081 AGSDVFDGDLGMAAGKLSLPHDPSPLQRYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1140
QY 1133 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVDVFAFGGAVENPEYLTPO 1192
```



Db 1141 NQPDVTRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVDVAFAGGAVENPEYLTPQ 1200  
QY 1193 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGDLVVP 1247  
Db 1201 GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGDLVVP 1255

## RESULT 10

AAG88267

ID AAG88267 standard; Protein; 1255 AA.

XX AAG88267;

DT 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;  
XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

XX WO200141787-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33591.

XX 10-DEC-1999; 99US-0458299.

XX (EPTM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
XX Keogh E;

XX WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
XX cellular immune responses for the prevention and treatment of cancer -  
XX Disclosure; Page 15; 199pp; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).  
XX Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
XX culture in vitro and binds to a complex of an epitope (I), bound to a  
XX human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising  
XX and a second epitope and the peptide is less than 50 contiguous amino  
XX acids that have 100% identity with a native peptide sequence of HER2/neu;  
XX (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
XX excipient; (4) an isolated nucleic acid encoding a peptide comprising  
XX (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic  
XX and immunostimulant activities, and can be used in vaccines. (I), (II)  
XX and (III) are useful for inducing cellular immune responses for the  
XX prevention and treatment of cancer. (I) and (II) are useful for  
XX monitoring or evaluating an immune response to a tumour-associated  
XX antigen when incubated with a T lymphocyte sample from a patient and  
XX detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
XX based vaccines mean that immunosuppressive epitopes that may be present  
XX in whole antigens may be avoided. Selected epitopes may be combined to  
XX enhance immunogenicity. The possible pathological side effects caused by  
XX infectious agents or whole protein antigen is eliminated. The vaccine  
XX provides the ability to direct and focus an immune response to multiple  
XX selected antigens from the same pathogen. Epitope-based anti-tumour  
XX vaccines provides the opportunity to combine epitopes derived from  
XX multiple tumour-associated molecules addressing the problem of tumour-  
XX tumour variability and reducing the likelihood of tumour escape due to  
XX antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
XX the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 22; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MELAALCRGLLALLPPGAASSTQVCTGTDMKRLRLPAS PETHLDMLRHLVGGCOVQGNL 60  
Db 1 MELAALCRGLLALLPPGAASSTQVCTGTDMKRLRLPAS PETHLDMLRHLVGGCOVQGNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVROVPLQRLRIVRGTLFEDNYALAVLDNG 120

QY 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNQOLA 180  
Db 121 DPLNNTTPTVGTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDITLWKDIFHKNQOLA 180

QY 181 LTLIDTNRSPACHPCSPMCKGSRGSESSDCOSLTRTVCAGGCARCKGPLPTDCCHQEC 240  
Db 181 LTLIDTNRSPACHPCSPMCKGSRGSESSDCOSLTRTVCAGGCARCKGPLPTDCCHQEC 240

QY 241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMOYIKANSKFIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVAVTSAN 360

QY 361 ELBFAGCKIRGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420  
Db 361 IOBFAGCKIRGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGYLYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWGLSLRSLRGLSGLALIHNTLHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWGLSLRSLRGLSGLALIHNTLHLCFVHTV 480

QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCGPPTQCVNCSQFLRQEC 540  
Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCGPPTQCVNCSQFLRQEC 540

QY 541 VVEECRVQLGPREYVNHARHCLPCHPECOPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600  
Db 541 VVEECRVQLGPREYVNHARHCLPCHPECOPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600

QY 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGCPEAORASPLTSTVSAVVG 660  
Db 601 PSGVKPDLSPYMPKPPDEEGACQPCPINCTHSCVDLDDKGCPEAORASPLTSTVSAVVG 660

QY 661 ILLVVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712  
Db 661 ILLVVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712

QY 713 RKVKVLGSGARGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 772  
Db 721 RKVKVLGSGARGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

QY 773 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGODLLNWCQIAKGSYLEDVR 832  
Db 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGODLLNWCQIAKGSYLEDVR 840

QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMMALESILRRRFT 892  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMMALESILRRRFT 900

QY 893 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKCWM 952  
Db 901 HQSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVYIMVKCWM 960

QY 953 IDSECRPRELVSPFSRMARDPQRFVWIONEDLGPASPLDSTFYSLLEDDMDGLVDA 1012  
Db 961 IDSECRPRELVSPFSRMARDPQRFVWIONEDLGPASPLDSTFYSLLEDDMDGLVDA 1020

QY 1013 EBYLVPQQGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLSEPEEAPRSLAPSEG 1072  
Db 1013 EBYLVPQQGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLSEPEEAPRSLAPSEG 1072

Db 1021 EELVLPQGGFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
QY 1073 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 1132  
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 1140  
QY 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1192  
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200  
QY 1193 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGPTTAENPEYLGDLVVP 1247  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSTFKGPTTAENPEYLGDLVVP 1255

## RESULT 11

AAE24067  
ID AAE24067 standard; Protein; 1255 AA.

XX AC AAE24067;

DT 23-SEP-2002 (first entry)

DE Human Her-2 protein.

XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;  
KW hyperproliferative disorder; prophylaxis; inflammation; antisense;  
KW tumour; gene therapy; phosphorothioate backbone.

XX Homo sapiens.

OS WO200222636-A1.

PN 21-MAR-2002.

PD 12-SEP-2001; 2001WO-US28572.

PF 15-SEP-2000; 2000US-0663834.

PR (ISIS-) ISIS PHARM INC.

PA Bennett CF, Cowsett LM;

PI WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

XX Novel antisense oligonucleotide which modulates the expression of Human  
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors  
PT inflammation or to prevent infection in humans -

XX Example 13; Page 95-107; 116pp; English.

XX The invention relates to antisense compounds targetted to a nucleic  
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)  
CC that specifically hybridises with and inhibits the expression of Her2.  
CC Antisense compounds of the invention are used for treating diseases or  
CC conditions associated with Her2 such as hyperproliferative disorders  
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,  
CC neural or cardiac cancer. They are also useful prophylactically e.g.  
CC to prevent or delay infection, inflammation and tumour formation. The  
CC invention is also used in gene therapy. The present sequence is human  
CC Her-2 protein.

XX Sequence 1255 AA;

QY Query Match 97.4%; Score 6602; DB 23; Length 1255;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLALLPPGAASVQCTGTDMLKRLPASPTHLDMLRHLVQGCQVQGNL 60  
Db 1 MELAALCRWGLLALLPPGAASVQCTGTDMLKRLPASPTHLDMLRHLVQGCQVQGNL 60

QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDI FHKNNOLA 180  
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDI FHKNNOLA 180  
QY 181 LTLIDNRSRACHPCSPMKCKSGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240  
Db 181 LTLIDNRSRACHPCSPMKCKSGRCWGESSEDCOSLTRTVCCAGCARCKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFITG 360  
Db 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360  
QY 361 ELEFACCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEEITGYLIYISAMPDLSL 420  
Db 361 IQEFACCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVFETLEEITGYLIYISAMPDLSL 420  
QY 421 DLSVFONLQVIRGRIILHNGAYSILTLOGLGISWLGSLRSLRGLSLALIHNTLHLCFVHTV 480  
Db 421 DLSVFONLQVIRGRIILHNGAYSILTLOGLGISWLGSLRSLRGLSLALIHNTLHLCFVHTV 480  
QY 481 PWDOLFNRNPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPGPTQCVNCSQPLRGOEC 540  
Db 481 PWDOLFNRNPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPGPTQCVNCSQPLRGOEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFEGPEADQCVACHYKDPFPCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECQONGSVTCFEGPEADQCVACHYKDPFPCVARC 600  
QY 601 PSGVKPDLSPYMPITWKFPDEBEGACQPCINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660  
Db 601 PSGVKPDLSPYMPITWKFPDEBEGACQPCINCTHSCVDLDDKCPAEQASPLTSTVSAVVG 660  
QY 661 ILLVVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETEL 712  
Db 661 ILLVVVLGVVFGILIKRRQKIRKYTMRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
QY 713 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 772  
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780  
QY 773 YVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSGODLLNWCNMQIAKMSYLEDVR 832  
Db 781 YVSRLLGICLTSTVQLVTOLMPYGCLLDHVRENRLGSGODLLNWCNMQIAKMSYLEDVR 840  
QY 833 LVHRDLAARNVLKSPNHNKIDTGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 892  
Db 841 LVHRDLAARNVLKSPNHNKIDTGLARLLDIDETEHADGGKVPKIMMALESILRRRFT 900  
QY 893 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPOLLEKGERLPDPPICTIDVYIMVWKCM 952  
Db 901 HQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPOLLEKGERLPDPPICTIDVYIMVWKCM 960  
QY 953 IDSECRPRFRELVSFERSWARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGDLVDA 1012  
Db 961 IDSECRPRFRELVSFERSWARDPQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGDLVDA 1020  
QY 1013 EELVFPQGGFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1072  
Db 1021 EELVFPQGGFFCDDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
QY 1073 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 1132  
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGVAPLTCSPQPEYV 1140  
QY 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1192

Db 1141 NOPDVRPQPPSPREGPLPAARAGATLERPKTSLSPCKNGVVKDVFAGGAVENPEYLTQ 1200  
QY 1193 CGAAPQHPHPPAFSPAFNLNYYWDDQPPRGAPPSTFKGTPTAENPEYLGLOVPV 1247  
Db 1201 CGAAPQHPHPPAFSPAFNLNYYWDDQPPRGAPPSTFKGTPTAENPEYLGLOVPV 1255

RESULT 12  
ID AAE20479  
XX AAE20479 standard; Protein; 1255 AA.  
AC AAE20479;  
DT 01-JUL-2002 (first entry)  
XX Human Her-2/neu protein.  
DE Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
XX Homo sapiens.

OS  
FH Key Location/Qualifiers  
FT Region 1021..1030  
FT /note= "Naturally processed HLA-B44-restricted epitope"  
FT  
FN WO200214503-A2.

PN 21-FEB-2002.  
XX 14-AUG-2001; 2001WO-US41733.  
XX 14-AUG-2000; 2000US-225152P.  
PR 28-SEP-2000; 2000US-236428P.  
PR 21-FEB-2001; 2001US-270520P.  
XX (CORI-) CORIXA CORP.  
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;  
PI Mcneill PD, Vedvick TS;  
XX WPI; 2002-280758/32.  
DR N-PSDB; AAD32743.  
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
PT prevention and diagnosis of cancer, preferably breast cancer  
PS Disclosure; Page 114-117; 129pp; English.

CC The invention relates to an isolated Her-2/Neu polypeptide composition  
CC effective for eliciting an immune response. The invention is useful for  
CC eliciting an immune response in a patient, where the patient is human  
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
CC The composition is useful for the therapy and diagnosis of cancer,  
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
CC and other compositions for the diagnosis, prevention and treatment of  
CC human malignancies, for stimulating and/or expanding T cells specific for  
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
CC patient. The invention is useful for stimulating a T cell response in a  
CC human patient, as probe or primer for nucleic acid hybridisation, to  
CC selectively form duplex molecules with complementary stretches of the  
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
CC length gene from a suitable library, and to direct expression of a  
CC polypeptide in appropriate host cells. The composition is useful in  
CC prophylactic or therapeutic applications and for the treatment of cancer,  
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
CC associated malignancies. The invention is useful in gene therapy. The  
CC present sequence is human Her-2/neu protein.

XX Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;

Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;  
QY 1 MELAALCRWGLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHLYOGCOVQGNL 60  
Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKLRLPASPETHLDMRLHLYOGCOVQGNL 60  
QY 61 ELYLPTNASLSFLQDIQEVQGYLIAHNQVRQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120  
Db 61 ELYLPTNASLSFLQDIQEVQGYLIAHNQVRQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120  
QY 121 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180  
Db 121 DFLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFHKNNQLA 180  
QY 181 LTLIDTNRSRACHPCSPCKGSRGSESSDCOSLTRTYCAGGCARCKGPLPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPCKGSRGSESSDCOSLTRTYCAGGCARCKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLNHNVTAEDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360  
QY 361 ELEFAGCKKIFGSLAFLPESDPDASNTAPLOEQLOVFETLEETITGYYISAWPDSLP 420  
Db 361 IQEAGCKKIFGSLAFLPESDPDASNTAPLOEQLOVFETLEETITGYYISAWPDSLP 420  
QY 421 DLSVFQNLQVIRGRIHNGAYSLTLQGLGISHWGLRSRLRELGLSLALIHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRIHNGAYSLTLQGLGISHWGLRSRLRELGLSLALIHNNTHLCFVHTV 480  
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFPCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKDPFPCVARC 600  
QY 601 PSGVKPDLSPYMPIWKFDEEGACQPCINCTHSCVDLDDKGPAPORASPLTSIYSAVVG 660  
Db 601 PSGVKPDLSPYMPIWKFDEEGACQPCINCTHSCVDLDDKGPAPORASPLTSIYSAVVG 660  
QY 661 ILLVVVLGVVFGILII---QYIKANS--KFIGITEL--PLTPSGAMPNQAQMRILKETEL 712  
Db 661 ILLVVVLGVVFGILIIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 713 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 772  
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGS 780  
QY 773 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNGLSGQDLNLCMQIAKMSYLEDVR 832  
Db 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNGLSGQDLNLCMQIAKMSYLEDVR 840  
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 892  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRFT 900  
QY 893 HQSDWVSYGVTWELMTFGAKPYDGIAREIPDLLEKGERLPDPPICTIDVYIMVKCWM 952  
Db 901 HQSDWVSYGVTWELMTFGAKPYDGIAREIPDLLEKGERLPDPPICTIDVYIMVKCWM 960  
QY 953 IDSECRPRFRELVSFSEFRMARDPQRFVVIQNEIDLGASPLDSTFYRSLEDDDDMGDLVDA 1012  
Db 961 IDSECRPRFRELVSFSEFRMARDPQRFVVIQNEIDLGASPLDSTFYRSLEDDDDMGDLVDA 1020  
QY 1013 EYLVPOGGFFCPDPAFGAGGMVHRHSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072

Db 1021 BEYLVPOQGFCCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG 1080  
Qy 1073 AGSDVFDGLGMAAKGLOSLTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1132  
Db 1081 AGSDVFDGLGMAAKGLOSLTHDPSPLOQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1133 NOPDVRPOPSPREGPLPAARAGATLERAKTSLPGKNGVWVDVFAFGGAVENPEYLTPO 1192  
Db 1141 NOPDVRPOPSPREGPLPAARAGATLERPKTSLPGKNGVWVDVFAFGGAVENPEYLTPO 1200  
Qy 1193 GGAARQPHPPAFSPAFDNLNLYWDQDPPRGAPPSTFKGTPTAENPEYLGDLVVPV 1247  
Db 1201 GGAARQPHPPAFSPAFDNLNLYWDQDPPRGAPPSTFKGTPTAENPEYLGDLVVPV 1255

RESULT 13  
AAMS1143  
ID AAMS1143 standard; Protein; 1255 AA.  
XX AC AAMS1143;  
XX DT 17-JUN-2002 (first entry)  
XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.  
XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;  
XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Domain 1..653 /note= "extracellular domain"  
FT Domain 676..1255 /note= "intracellular domain"  
FT Domain 990..1255 /note= "phosphorylation domain"  
XX FT  
XX WO200212341-A2.  
XX PN  
XX PD 14-FEB-2002.  
XX PF 03-AUG-2001; 2001WO-US24283.  
XX PR 03-AUG-2000; 2000US-0632507.  
XX PA (CORI-) CORIXA CORP.  
XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PI Cheever MA, Gheyssen D;  
XX DR WPI; 2002-241743/29.  
XX DR N-PSDB; ABA92250.  
XX PT Her-2/neu fusion protein for treating or preventing cancer by eliciting  
PT or enhancing an immune response to the protein, has Her-2/neu  
PT extracellular domain fused to Her-2/neu intracellular or  
PT phosphorylation domain -  
XX PS Claim 68; Fig 7; 14lpp; English.  
XX CC The present sequence is that of human Her-2/neu (p185 glycoprotein  
CC or c-erbB2), an oncogenic self-protein and target for anti-cancer  
CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed  
CC in a variety of cancers, including breast, ovarian, colon, lung and  
CC prostate cancer. Her-2/neu is a member of the tyrosine kinase  
CC family of receptor-like glycoproteins. It comprises an extracellular  
CC domain with homology to the epidermal growth factor receptor  
CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal  
CC intracellular domain that also shows homology to EGFR. Its  
CC overexpression correlates with a poor prognosis in breast and  
CC ovarian cancers. The invention provides Her-2/neu fusion  
CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In  
CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
CC protein is fused to a Her-2/neu intracellular domain or  
CC phosphorylation domain (or its DeltaPD fragment). An immune  
CC response to Her-2/neu protein is elicited or enhanced by  
CC administering the fusion protein in the form of a vaccine, or by  
CC transfusing cells of an animal ex vivo with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.4%; Score 6602; DB 23; Length 1255;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;  
Qy 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHLYQGCVVQGNL 60  
Qy 61 ELTYLPTNASLSFLODIOEVQGYVLIHNVQVPLQRLRIVRGTLQDFEDNALAVLDNG 120  
Db 61 ELTYLPTNASLSFLODIOEVQGYVLIHNVQVPLQRLRIVRGTLQDFEDNALAVLDNG 120  
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQPNPQLCYQDTILMKDIFHKKNQLA 180  
Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQPNPQLCYQDTILMKDIFHKKNQLA 180  
Qy 181 LTLIDNRSRACHPCSPCKRCWGSESDCOSLTRTVCAGCARCKGPLPTDCCHEQC 240  
Db 181 LTLIDNRSRACHPCSPCKRCWGSESDCOSLTRTVCAGCARCKGPLPTDCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
-Y 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMEHLREVRVTSAN 360  
Qy 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFLEETITGYLYISAWPDSL 420  
Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFLEETITGYLYISAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLGLAIHHNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLSLRELGLGLAIHHNTHLCFVHTV 480  
Qy 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCNCVSQFLRGQEC 540  
Db 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPGTQCNCVSQFLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHYKDPFPCVARC 600  
Qy 601 PSGVKPDLISYMPIMKFPDEEGACQPCPNCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660  
Db 601 PSGVKPDLISYMPIMKFPDEEGACQPCPNCTHSCVDLDDKGCPEARASPLTSIVSAVVG 660  
Qy 661 ILLVVLGVVFGILII---QYIKANS--KFIGITEL--PLTPSGAMPNQAQRILKETEL 712  
Db 661 ILLVVLGVVFGILIIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQRILKETEL 720  
Qy 713 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 772  
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

QY 773 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDQLLWNCQIAKMSYLEVDVR 832  
DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDQLLWNCQIAKMSYLEVDVR 840  
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 892  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
QY 893 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCM 952  
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCM 960  
QY 953 IDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1012  
DB 961 IDSECRPRFRELSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA 1020  
QY 1013 EYLVPQOGEFCDDPAPGAGGWHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072  
DB 1021 EYLVPQOGEFCDDPAPGAGGWHHRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
QY 1073 AGSDVFDGDLGMAAGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1132  
DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1133 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1192  
DB 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
QY 1193 GGAAPQHPHPPAFSPAEDNLYYWDODPPERGAPPSTFKGTPTAENPEYLGLDVDPV 1247  
DB 1201 GGAAPQHPHPPAFSPAEDNLYYWDODPPERGAPPSTFKGTPTAENPEYLGLDVDPV 1255

## RESULT 14

AAU77114

ID AAU77114 standard; Protein; 1255 AA.

XX AC AAU77114;

XX XX

XX 05-JUN-2002 (first entry)

XX Human Her-2/neu polypeptide.

XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;

XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;

XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;

XX Hodgkin's lymphoma; T cell therapy.

XX Homo sapiens.

XX WO200213847-A2.

XX 21-FEB-2002.

XX 13-AUG-2001; 2001WO-US25408.

XX 14-AUG-2000; 2000US-0638280.

XX 28-SEP-2000; 2000US-0675904.

XX (CORI-) CORIXA CORP.

XX Gaiger A, Cheever MA, Hand-zimmermann S;

XX WPI; 2002-280741/32.

XX N-PSDB; ABK10730.

XX Inhibiting haematological malignancy development by administering

XX polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide

XX PT encoding the polypeptide, or antigen presenting cells expressing the

XX polypeptide

XX Disclosure; Page 71-74; 74pp; English.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC The invention relates to a method for inhibiting development of  
CC haematological malignancy in a patient by administering a polypeptide  
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
CC encoding the polypeptide. Antigen presenting cells that express the  
CC protein can also be administered. The sequences are used for inhibiting  
CC development of haematological malignancy such as acute myelogenous  
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
CC leukaemia (CLL), MDS, myeloma, Hodgkin's lymphoma and non-Hodgkin's  
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6602; DB 23; Length 1255;  
Best Local Similarity 97.2%; Pred. No. 0;  
Matches 1220; Conservative 9; Mismatches 18; Indels 8; Gaps 3;

QY 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMKRLPASPETHLDMLRHLYGCCVQVQGNL 60  
DB 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMKRLPASPETHLDMLRHLYGCCVQVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLIRVIRGTQLPEDNYALAVLNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLQRLIRVIRGTQLPEDNYALAVLNG 120  
QY 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180  
DB 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGSRGWGSESDCQSLTRTVACAGGCARCKPLTDCCHQC 240  
DB 181 LTLIDTNRSRACHPCSPMKGSRGWGSESDCQSLTRTVACAGGCARCKPLTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
QY 361 ELSEFAGCKITFGLAFIPESFDGDPASNTAPLQPEQLQVFTLEETIGLYLISAWPDSLP 420  
DB 361 IQSEFAGCKITFGLAFIPESFDGDPASNTAPLQPEQLQVFTLEETIGLYLISAWPDSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSLRELGLSLALHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRLSLRELGLSLALHNNTHLCFVHTV 480  
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTCVNCSCQFLRGQEC 540  
DB 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTCVNCSCQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600  
DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600  
QY 601 PSYGKPDLSYMPIWKPPDEGACQPCPINCTHSCVDLDDKGCPCPAEQASPLTISVAVVG 660  
DB 601 PSYGKPDLSYMPIWKPPDEGACQPCPINCTHSCVDLDDKGCPCPAEQASPLTISVAVVG 660  
QY 661 ILLVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETTEL 712  
DB 661 ILLVVLGVVFGILI---QYIKANS--KFIGITEL--PLTPSGAMPNQAMRILKETTEL 712  
QY 713 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVNAVGVSP 772  
DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVNAVGVSP 780  
QY 773 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDQLLWNCQIAKMSYLEVDVR 832  
DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDQLLWNCQIAKMSYLEVDVR 840  
QY 833 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 892

Db 841 LVHRDLARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPKMKMALESILRRFT 900  
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVKWM 952  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960  
Qy 953 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1012  
Db 961 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1020  
Qy 1013 EYLVLPQOGFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072  
Db 1021 EYLVLPQOGFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
Qy 1073 AGSDVFDGDLGMGAAGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBEYV 1132  
Db 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBEYV 1140  
Qy 1133 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVVDVFAFGAVENPEYLTPO 1192  
Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVVVDVFAFGAVENPEYLTPO 1200  
Qy 1193 GGAAPQHPHPPAFSPADNLVYWDQDPPERGAFTSTFKGTPTABNPEYLGIDVPV 1247  
Db 1201 GGAAPQHPHPPAFSPADNLVYWDQDPPERGAFTSTFKGTPTABNPEYLGIDVPV 1255

## RESULT 15

AAR39568  
ID AAR39568 standard; Protein; 1433 AA.  
AC AAR39568;  
XX  
DT 07-FEB-1994 (first entry)  
XX  
DE Sequence of c-erbB-2 tumour antigen.  
XX  
KW Tumour antigen; c-erbB-2; glycoprotein.  
XX  
OS Homo sapiens.  
XX  
FN WO9316185-A.  
XX  
PD 19-AUG-1993.  
XX  
PF 05-FEB-1993; 93WO-US01055.  
XX  
PR 06-FEB-1992; 92US-0831967.  
XX  
PA (CETU ) CETUS ONCOLOGY CORP.  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
PI Houston Lb, Huston JS, Oppermann H, Ring DB;  
XX  
DR WPI; 1993-272889/34.  
DR N-PSDB; AAQ46083.  
XX  
PT New single chain Fv polypeptide binding to C-erbB-2 tumour  
PT antigen - for imaging or treating breast or ovarian cancer etc.  
XX  
PS Disclosure; pages 48-54; 87pp; English.  
XX  
CC c-erbB-2 refers to a protein antigen expressed on the surface of  
CC tumour cells, such as breast and ovarian tumour cells, which is an  
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents  
CC the location of a stop codon in AAQ46083.  
XX  
SQ Sequence 1433 AA;

Query Match 96.8%; Score 6559; DB 14; Length 1433;  
Best Local Similarity 96.7%; Pred. No. 0;

Matches 1213; Conservative 11; Mismatches 23; Indels 8; Gaps 3;  
Qy 1 MELAALCRWGLLLALLPFGAASCTGCTOMKRLPASPETHLDMKRLHYGCCVQVQGNL 60  
Db 1 MELAALCRWGLLLALLPFGAASCTGCTOMKRLPASPETHLDMKRLHYGCCVQVQGNL 60  
Qy 61 ELTYLPTNASISFLQDIOEVQYVLI AHNOVRQVPLORLRIVRGTPQLPFDNVALAVLNG 120  
Db 61 ELTYLPTNASISFLQDIOEVQYVLI AHNOVRQVPLORLRIVRGTPQLPFDNVALAVLNG 120  
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPQLCYQDTILWKDILFKHKNQLA 180  
Db 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPQLCYQDTILWKDILFKHKNQLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKGSRGCSWESSEDCQSLTRTVCAAGGACRCKGPLPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGCSWESSEDCQSLTRTVCAAGGACRCKGPLPTDCCHQC 240  
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLLVCLPHNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLLVCLPHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVAVTSAN 360  
Qy 361 ELBPAGCKITFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLBEITCYLYISAMPDSL 420  
Db 361 IQEPAGCKITFGSLAFIPESFDGDPASNTAPLQPEHLQVFTLEQITCYLYISAMPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTGLGTSWGLRSRLGSLGSLALHNNTHLFCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTGLGTSWGLRSRLGSLGSLALHNNTHLFCFVHTV 480  
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQCHARGHGWPGPTQCNCVSQFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQCHARGHGWPGPTQCNCVSQFLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
Qy 601 PSGVKPDLSTYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPABORASPLTSIVS 660  
Db 601 PSGVKPDLSTYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGCAPABORASPLTSIVS 660  
Qy 661 ILLVVVLGVVFGILI ---QYIKANS--KFIGITEL--PLTPSGAMPNQAOHRIILKETEL 712  
Db 661 ILLVVVLGVVFGILI KRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAOHRIILKETEL 720  
Qy 713 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 772  
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Qy 773 YVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRGLSGDLLNWCMIAGMSYLEVDV 832  
Db 781 YVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRGLSGDLLNWCMIAGMSYLEVDV 840  
Qy 833 LVHRDLARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPKMKMALESILRRFT 892  
Db 841 LVHRDLARNVVKSPNHVKITDFGLARLLDIDETEHADGCKVPKMKMALESILRRFT 900  
Qy 893 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVKWM 952  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYIMVKWM 960  
Qy 953 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1012  
Db 961 IDSECRPRFELVSEFSRMDRPFVVIQNEDELGPASPLDSTFYRSLLDDMDGLVDA 1020  
Qy 1013 EYLVLPQOGFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1072  
Db 1021 EYLVLPQOGFFCPDPAFCAGGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080

Qy	1073	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDTVPLPSETDGYVAPLTCSPQPEYV	1132
Db	1081	AGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1133	NOPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO	1192
Db	1141	NOPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
Qy	1193	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1247
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 09:16:33  
Job time : 41.6984 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds  
(without alignments)  
4403.399 Million cell updates/sec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6657	97.9	1255	21	Human heregulin 2
2	6657	97.9	1255	22	Human tyrosine kin
3	6657	97.9	1255	22	HER2 transgene pla
4	6657	97.9	1255	23	Human HER2 (ErbB2)
5	6651	97.8	1255	17	AAW01111
6	6651	97.8	1255	20	AAW92406
7	6651	97.8	1255	21	AAW921198
8	6651	97.8	1255	21	Human HER-2/neu pr
9	6651	97.8	1255	22	Amino acid sequenc
10	6651	97.8	1255	22	Human HER-2/neu pr

11	6651	97.8	1255	23	AAE24067	Human Her-2 protei
12	6651	97.8	1255	23	AAE20479	Human Her-2/neu pr
13	6651	97.8	1255	23	AAE51143	Human Her-2/neu on
14	6651	97.8	1255	23	AAU77114	Human Her-2/neu po
15	6651	97.8	1255	23	AAU77114	Sequence of c-erbB
16	6651	97.8	1255	23	AAU77114	Human breast cance
17	6651	97.8	1255	23	AAU77114	Human HER-2/neu pr
18	6651	97.8	1255	23	AAU77114	Rat Her-2/neu prot
19	6651	97.8	1255	23	AAU77114	Human HER-2/neu pr
20	6651	97.8	1255	23	AAU77114	Mouse Her-2/neu pr
21	6651	97.8	1255	23	AAU77114	Amino acid sequenc
22	6651	97.8	1255	23	AAU77114	Mouse Her-2/neu on
23	6651	97.8	1255	23	AAU77114	Human HER-2/neu fu
24	6651	97.8	1255	23	AAU77114	Her-2/neu extracel
25	6651	97.8	1255	23	AAU77114	Mouse Her-2/neu ex
26	6651	97.8	1255	23	AAU77114	Mouse Her-2/neu ex
27	6651	97.8	1255	23	AAU77114	Human HER-2/neu fu
28	6651	97.8	1255	23	AAU77114	Her-2/neu extracel
29	6651	97.8	1255	23	AAU77114	Her2-GM-CSF immuno
30	6651	97.8	1255	23	AAU77114	Extracellular HER-
31	6651	97.8	1255	23	AAU77114	Human Her-2/neu on
32	6651	97.8	1255	23	AAU77114	Human ErbB2 oncopr
33	6651	97.8	1255	23	AAU77114	Human ErbB2 extrac
34	6651	97.8	1255	23	AAU77114	DC9scFv-erbB2EC fu
35	6651	97.8	1255	23	AAU77114	Extracellular port
36	6651	97.8	1255	23	AAU77114	Amino acid sequenc
37	6651	97.8	1255	23	AAU77114	Human EGF receptor
38	6651	97.8	1255	23	AAU77114	Human Her-1 protei
39	6651	97.8	1255	23	AAU77114	Human epidermal gr
40	6651	97.8	1255	23	AAU77114	Amino acid sequenc
41	6651	97.8	1255	23	AAU77114	Human protein for
42	6651	97.8	1255	23	AAU77114	Human protein for
43	6651	97.8	1255	23	AAU77114	Human protein for
44	6651	97.8	1255	23	AAU77114	Human protein for
45	6651	97.8	1255	23	AAU77114	Human epidermal gr

#### ALIGNMENTS

RESULT 1  
ID AAY92620 standard; Protein; 1255 AA.  
AC AAY92620;  
DT 10-AUG-2000 (first entry)  
XX Human heregulin 2 (Her2).  
DE Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
KW self-protein; cancer; breast cancer; prostate cancer;  
KW cell-associated peptide antigen; foreign epitope.  
XX Homo sapiens.  
OS Homo sapiens.  
FH Key  
FH Domain  
FT Location/Qualifiers  
FT 1..173  
FT /label= N-terminal  
FT /note= "mature polypeptide"  
FT 5..25  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 59..73  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 103..117  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 149..163  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 174..323  
FT Domain



FT Region /label= Cysteine\_rich\_domain  
 FT 210..224 /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 250..264  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 324..483  
 FT /label= Ligand\_binding\_domain  
 FT 325..339  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 369..383  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 465..479  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 484..623  
 FT /label= Cysteine\_rich\_domain  
 FT 579..593  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 624..654  
 FT /label= Transmembrane\_domain  
 FT 632..652  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 653..667  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 655..1010  
 FT /label= Tyrosine\_kinase\_domain  
 FT 661..675  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 695..709  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 710..730  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 1011..1235  
 FT /label= C-terminal\_domain  
 XX WO200020027-A2.  
 XX 13-APR-2000.  
 XX 05-OCT-1999; 99WO-DK00525.  
 XX 05-OCT-1998; 98DK-0001261.  
 XX 20-OCT-1998; 98US-0105011.  
 XX (MEBI-) M & E BIOTECH AS.  
 XX Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
 XX Gautam A, Birk P, Karlsson G;  
 XX WPI; 2000-349917/30.  
 XX N-PSDB; AAA09455.  
 XX Inducing immune responses to weakly immunogenic, tumor associated  
 XX peptide antigens for the treatment of breast and prostate cancer  
 XX Claim 62; Page 193-198; 220pp; English.  
 XX This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
 XX Her2 can be used in the claimed method as an autovaccine to induce a CTL  
 XX response. Subdominant CTL epitopes, antibody binding regions and  
 XX cysteine residues involved in disulfide bonds are preserved in the  
 XX immunogenized forms. Regions suitable for the insertion of foreign T  
 XX helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
 CC cell-associated peptide antigens (PA) such as those associated with  
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
 CC The method comprises effecting simultaneous presentation by antigen  
 CC producing cells (APCs) of the animals immune system of: (1) at least 1  
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
 CC B-cell group derived from the cell-associated PA; and (2) at least 1  
 CC first T helper cell group which is foreign to the animal. Analogues of  
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
 CC part of all known and predicted CTL and B-cell epitopes of the respective  
 CC PA and including at least one foreign T helper epitope are also claimed.  
 CC The method is used to treat prostate, prostate/breast or breast cancer  
 CC when the PA is human PSM, FGF8b and Her2, respectively.  
 XX  
 XX Sequence 1255 AA;  
 Query Match 97.9%; Score 6657; DB 21; Length 1255;  
 Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 1227; Conservative 9; Mismatches 19; Indels 0; Gaps 0;  
 QY 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60  
 DB 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60  
 QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVQVPLQRLRIVRGTLFEDNALVALVDNG 120  
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVQVPLQRLRIVRGTLFEDNALVALVDNG 120  
 QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQPNQLCYQDTILWKDIFHKNNQLA 180  
 DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQPNQLCYQDTILWKDIFHKNNQLA 180  
 QY 181 LTLIDTNRSRACHPCSPMCKSRGWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240  
 DB 181 LTLIDTNRSRACHPCSPMCKSRGWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSPCARVCYGLGMQVYIKANSKFIGIT 360  
 DB 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSPCARVCYGLGMHLREVRVTSAN 360  
 QY 361 ELEFAGCKKIFGSLAPLPESPDGDPASNTAPLOPEOLQVFETLEEITGYLYISAWPDSL 420  
 DB 361 IOEFAGCKKIFGSLAPLPESPDGDPASNTAPLOPEOLQVFETLEEITGYLYISAWPDSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLGLALIHNNTHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLGLALIHNNTHLCFVHTV 480  
 QY 481 PWDQLFRNPQHALHTANRPEDECVGEGLAHQLCARGHCWPGPTQCNCVSQFLRGQEC 540  
 DB 481 PWDQLFRNPQHALHTANRPEDECVGEGLAHQLCARGHCWPGPTQCNCVSQFLRGQEC 540  
 QY 541 VEECRVLQGLPREYVNHARCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFPCVARC 600  
 DB 541 VEECRVLQGLPREYVNHARCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFPCVARC 600  
 QY 601 PQYIKANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660  
 DB 601 PSGVFPDLSJYMPFWKFPDEEGACQPCPINCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660  
 QY 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720  
 DB 661 ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAQRILKETEL 720  
 QY 721 RKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYMAGVGS 780  
 DB 721 RKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYMAGVGS 780



```
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Dy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRLGSGDQLLNWCQIAKMSYLEDVR 840
Dy 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRLGSGDQLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWMALESILRRFT 900
Dy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWMALESILRRFT 900
Qy 901 HQSDWMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Dy 901 HQSDWMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Qy 961 IDSECRPRFRELVSFMSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Dy 961 IDSECRPRFRELVSFMSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLVDA 1020
Qy 1021 EYLVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLPSEBEAPRSLAPSEG 1080
Dy 1021 EYLVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLPSEBEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Dy 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPLOYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQPDVRQPPSPREGPLPAAPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Dy 1141 NQPDVRQPPSPREGPLPAAPAGATLERAKTLSPGKNGVVDVFAFGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGIDVPV 1255
Dy 1201 GGAAPQHPHPPAFSPFDNLVYWDQDPPERCAPPSTFKGTPTAENPEYLGIDVPV 1255
```

## RESULT 3

```
ID AAB60167 standard; Protein; 1255 AA.
XX
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
XX antibody.
XX Homo sapiens.
XX Synthetic.
XX OS
XX PN W0200100244-A2.
XX
PD 04-JAN-2001.
XX
PF 23-JUN-2000; 2000WO-US17229.
XX
PR 25-JUN-1999; 99US-0141316.
XX
PR 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
XX Erickson S, Schwall R;
XX
XX WPI; 2001-061962/07.
XX
XX N-PSDB; AAF24297.
XX
```

Treating tumors, particularly breast cancers, which overexpress an ErbB receptor and does not respond to an anti-ErbB antibody, comprises conjugating the antibody to a maytansinoid -

## Example 3; Fig 4; 92pp; English.

The present invention provides a method of treating cancer by administering a conjugate of anti-ErbB antibody with a maytansinoid. In particular, the antibody is directed against ErbB2 (also known as HER2 and p185neu). The method is particularly useful in the treatment of breast, ovarian, stomach, endometrial, salivary gland, lung, kidney, colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

Seq Sequence 1255 AA;

Query Match 97.9%; Score 6657; DB 22; Length 1255;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 1227; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

```
Qy 1 MELAALCRWGLLLALLPPGAASQTCTGDMKLRLPASPETHLDMRLHYGQCQVQGNL 60
Dy 1 MELAALCRWGLLLALLPPGAASQTCTGDMKLRLPASPETHLDMRLHYGQCQVQGNL 60
Qy 61 ELTYLPTNASLFLQDIQEVGYVLIAHNVQVPLQRLRIVRGTLQFEDNVYALAVLNG 120
Dy 61 ELTYLPTNASLFLQDIQEVGYVLIAHNVQVPLQRLRIVRGTLQFEDNVYALAVLNG 120
Qy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQIRNPQLCYQDTILWKDIFHKNNOLA 180
Dy 121 DPLNNTTPTVTGASPGGLRELQRLSLTEILKGVLIQIRNPQLCYQDTILWKDIFHKNNOLA 180
Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCCAGGCARCKGPLPTDCCHQEC 240
Dy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCCAGGCARCKGPLPTDCCHQEC 240
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYRTFGASCVTACP 300
Dy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGYRTFGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360
Dy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHREVRVTSAN 360
Qy 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFTLEBEITGLYISAWPDSLP 420
Dy 361 IQEFAGCKKIFGSLAFLPESPDGPASNTAPLQPEQLQVFTLEBEITGLYISAWPDSLP 420
Qy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGSIWLGRLSRLRELGSLALHNNHLCFVHTV 480
Dy 421 DLSVFQNLQVIRGRILHNGAYSLSLTQGLGSIWLGRLSRLRELGSLALHNNHLCFVHTV 480
Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGEGACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Dy 481 PWDQLFRNPHQALLHTANRPEDECVEGEGACHQLCARGHCWGPGPTQCVNCSQFLRGQEC 540
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCVACAHYKDPFPCVARC 600
Dy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCVACAHYKDPFPCVARC 600
Qy 601 POYIKANSKFITIGITELPDEEGACOPCPINCTHSCVDLDDKGCPCAEORASPLTSIVSAVVG 660
Dy 601 PSYVPLSYWPIWKPFDEEGACOPCPINCTHSCVDLDDKGCPCAEORASPLTSIVSAVVG 660
Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLEQETELVEPLTPSGAMPNQAQMRILKETEL 720
Dy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRLLEQETELVEPLTPSGAMPNQAQMRILKETEL 720
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Dy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRLGSGDQLLNWCQIAKMSYLEDVR 840
Dy 781 YVSRLLGICLTSTVOLVTQMLPYGCLLDHVRENRLGSGDQLLNWCQIAKMSYLEDVR 840
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWMALESILRRFT 900
Dy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWMALESILRRFT 900
```

Db 841 LVHRDLARNVLKSPNVHKITDFGLARLLDIDETEHADGKGVPIKWMALLESILRRFT 900  
 Qy 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVTYIMVVKWM 960  
 Db 901 HQSDVMSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVTYIMVVKWM 960  
 Qy 961 IDSECRPFRELVSFSSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA 1020  
 Db 961 IDSECRPFRELVSFSSRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDMDGLVDA 1020  
 Qy 1021 EBYLVPOQGFCCPDAPGAGGVMVHRHRSSTRSGGDLTLCLEPSEEEAPRSLAPSEG 1080  
 Db 1021 EBYLVPOQGFCCPDAPGAGGVMVHRHRSSTRSGGDLTLCLEPSEEEAPRSLAPSEG 1080  
 Qy 1081 AGSDVFDGDLGMAAKGLQSLTPHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQRYV 1140  
 Db 1081 AGSDVFDGDLGMAAKGLQSLTPHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQRYV 1140  
 Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAFGAVENPEYLTPO 1200  
 Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAFGAVENPEYLTPO 1200  
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVPU 1255  
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLGLDVPU 1255

## RESULT 4

AAU74545  
 ID AAU74545 standard; Protein; 1255 AA.  
 XX  
 AC AAU74545;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human HER2 (ErbB2) polypeptide.  
 XX

KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glial disorder; astrocytal disorder; hypothalamic disorder;  
 KW glandular disorder; macrophagal disorder; epithelial disorder;  
 KW stromal disorder; blastocoealic disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-0811123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 97.9%; Score 6657; DB 23; Length 1255;  
 Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 1227; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTVCTGDMKRLRPASPETHDMLRLHYOGCVVQGNL 60  
 Db 1 MELAALCRWGLLLALLPPGAASTVCTGDMKRLRPASPETHDMLRLHYOGCVVQGNL 60  
 Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNVALAVLNG 120  
 Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTLQFEDNVALAVLNG 120  
 Qy 121 DPLNTPPTVTCASPGGLRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
 Db 121 DPLNTPPTVTCASPGGLRELQRLSITELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
 Qy 181 LTLIDTNRSRACHPCSPMCKSGCWGESSEDCQSILTRTVACGGCARGKPLTDCCHEOC 240  
 Db 181 LTLIDTNRSRACHPCSPMCKSGCWGESSEDCQSILTRTVACGGCARGKPLTDCCHEOC 240  
 Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Qy 301 YNYLSTDVGSCTLVCPILHNQVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGIT 360  
 Db 301 YNYLSTDVGSCTLVCPILHNQVTAEDGTQRCCKSKPCARVCYGLGMHLEVRVAVTSAN 360  
 Qy 361 ELEFAGCKKIFGSLAFLPESPDGDPASNTAPLOEQLOVFTLEITCYLYISAMPDSL 420  
 Db 361 IOEFAGCKKIFGSLAFLPESPDGDPASNTAPLOEQLOVFTLEITCYLYISAMPDSL 420  
 Qy 421 DLSVFQNLQVIRGIRILHNGAYSLTLQGLIGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGIRILHNGAYSLTLQGLIGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480  
 Qy 481 PWDLFNPHQALLHTANRPEDECVGEGGLACHQICARGHCGWPGPTQCVNCSQFLRGQEC 540  
 Db 481 PWDLFNPHQALLHTANRPEDECVGEGGLACHQICARGHCGWPGPTQCVNCSQFLRGQEC 540  
 Qy 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGPEADQCVACAHYKDPFPCVARC 600  
 Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGPEADQCVACAHYKDPFPCVARC 600  
 Qy 601 PQVIKANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTSTVSAVVG 660  
 Db 601 PSGVKPDLSYMPWKPFDEEGACQPCPINCTHSCVDLDDKGCPCAEORASPLTSTVSAVVG 660  
 Qy 661 ILLVVLGVVFGILIKRRQKIRKYTMRLLEQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Db 661 ILLVVLGVVFGILIKRRQKIRKYTMRLLEQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Qy 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVNAVGVSP 780  
 Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVNAVGVSP 780  
 Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNENRGRIGSQDLLNWCNQIAGKHSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHRVNRGRGLSQDILLNWCMIKAGMSYLEDVR 840  
QY LVHRDLAARNVLVAKPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALSIILRRRT 900  
Db 841 LVHRDLAARNVLVAKPNHVKITDFGLARLLDIDETEHADGCKVPIKWMALSIILRRRT 900  
QY HQSDVMSYGVTVWELMTFCAKPYDGI PAREIPDLLEKGERLPQPPICTIDVTYMIWVKWM 960  
Db 901 HQSDVMSYGVTVWELMTFCAKPYDGI PAREIPDLLEKGERLPQPPICTIDVTYMIWVKWM 960  
QY 961 IDSECRPRELVSEFSRMRDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
Db 961 IDSECRPRELVSEFSRMRDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
QY 1021 EYLVVPOQGFCCPDAPAGAGGVMHRRSSSTRSGGDLTLGLEPSEERAPRSLAPSG 1080  
Db 1021 EYLVVPOQGFCCPDAPAGAGGVMHRRSSSTRSGGDLTLGLEPSEERAPRSLAPSG 1080  
QY 1081 AGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140  
Db 1081 AGSDVFDGDLGMAAGKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEYV 1140  
QY 1141 NQPDVVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVWVDVFAFGAVENPEYLTPO 1200  
Db 1141 NQPDVVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVWVDVFAFGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLOVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKTPTAENPEYLGLOVPV 1255

RESULT 5

AAW01111

ID AAW01111 standard; Protein; 1255 AA.

XX AA

AC AAW01111;

DT 01-JAN-1997 (first entry)

XX HER-2/neu protein.

DE HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;

KW breast cancer; ovary cancer; colon cancer; lung cancer;

KW prostate cancer; immunisation; tumour; vaccine; vector.

XX Homo sapiens.

XX Location/Qualifiers

FT Domain 676..1255

FT /label= Intracellular domain

FT /note= "claimed domain, useful for immunisation"

XX WO9630514-A1.

XX 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNITW ) UNIV WASHINGTON.

XX Cheever MA, Disis ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or

XX treatment of malignancies with which the HER-2/neu oncogene is

XX associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.

XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 17; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPETHLDMLRHLYQGCQVVGNNL 60

Db 1 MELAALCRWGLLLALLPPGAASQTCTGTDMLRLPASPETHLDMLRHLYQGCQVVGNNL 60

QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRI VRGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVROVPLQRLRI VRGTQLFEDNYALAVLDNG 120

QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180

QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARKGKPLPTDCCHEQC 240

Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARKGKPLPTDCCHEQC 240

QY 241 RAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPREGRTFCASCVTACP 300

Db 241 RAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESNPREGRTFCASCVTACP 300

QY 301 NYNLSTDVSGSCTLVCPHLNQEVTAEDGTQRCCKSPKPCARVCYGLGMQVIKANSKFIGIT 360

Db 301 NYNLSTDVSGSCTLVCPHLNQEVTAEDGTQRCCKSPKPCARVCYGLGMQVIKANSKFIGIT 360

QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAMPDLSLP 420

Db 361 IQEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEQLQVFETLEEITGYLYISAMPDLSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGURSRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGURSRELGSGLALIHNNTHLCFVHTV 480

QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540

Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC 540

QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVARC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCVARC 600

QY 601 POYIKANSKFITELPDEEGACQPCPNCTHSCVDLDDKGCPCARASPLTSISAVVG 660

Db 601 PSGVKPDLSYMPDWKPFDEEGACQPCPNCTHSCVDLDDKGCPCARASPLTSISAVVG 660

QY 661 ILLVVVLGVVFGILIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAQRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRRQKIRKYTMRLRLQETELVEPLTPSGAMPNQAQRILKETEL 720

QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGS 780

Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGS 780

QY 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHRVNRGRGLSQDILLNWCMIKAGMSYLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTLQMPYGCLLDHRVNRGRGLSQDILLNWCMIKAGMSYLEDVR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960  
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960  
QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDGDLVDA 1020  
DB 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEDELGPASPLDSTFYRSLLEDGDLVDA 1020  
QY 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080  
DB 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
DB 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPPPAPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVAV 1255  
DB 1201 GGAAPQHPPPAPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVAV 1255

## RESULT 6

ID AAW92406  
XX AAW92406 standard; Protein; 1255 AA.  
AC AAW92406;  
XX 21-APR-1999 (first entry)  
XX Human HER-2/neu oncogene protein.  
XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
XX malignancy; treatment; tumour.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX Region 676..1255  
XX /note= "region which elicits immune response"

XX US5869445-A.  
XX 09-FEB-1999.  
XX 01-APR-1996; 96US-0625101.  
XX 01-APR-1996; 96US-0625101.  
XX 17-MAR-1993; 93US-0033644.  
XX 12-AUG-1993; 93US-0106112.  
XX 31-MAR-1995; 95US-0414417.  
XX (UNIW ) UNIV WASHINGTON.  
XX Cheever MA, Disis ML;  
XX WPI; 1999-152835/13.  
XX N-PSDB; AAX01912.

XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
XX an HER-2/neu associated malignancy, particularly for treating or  
XX preventing tumours  
XX Claim 3; Column 31-38; 26pp; English.  
XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune  
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
CC B cells to produce an immune response to the HER-2/neu protein. The  
CC method can be used for immunisation against a malignancy in which the  
CC HER-2/neu oncogene is associated and in the treatment of an existing  
CC tumour, or to prevent tumour occurrence or recurrence.

XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 20; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60  
DB 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60  
C/ 61 ELTYLPTNASLSFLQDIOEVQYVLIHNNQVQVPLQRLIRVGTQLPEDNYALAVLNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQYVLIHNNQVQVPLQRLIRVGTQLPEDNYALAVLNG 120  
QY 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDI FHKNNOLA 180  
DB 121 DPLNNTTPTVGTASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDI FHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGSCWGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQC 240  
DB 181 LTLIDTNRSRACHPCSPMKGSCWGESSEDCQSLTRTVCAAGCARCKGPLTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSLVCPPLHNOEVTAEADGTQCEKSKPCARVCYGLGHQYIKANSKFIGIT 360  
DB 301 YNYLSTDVGSLVCPPLHNOEVTAEADGTQCEKSKPCARVCYGLGHQYIKANSKFIGIT 360  
QY 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFETLEEITGLYISAMPDLSL 420  
DB 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFETLEEITGLYISAMPDLSL 420  
QY 421 DLSVFQNLQVIRGRIHNGAYSLSLQGLGISWGLSLRSLRELGLALIHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRIHNGAYSLSLQGLGISWGLSLRSLRELGLALIHNNTHLCFVHTV 480  
QY 481 PWQLFRNPHOALLHTANRPEDECVGGLACHOL-CARGHCGPGTCCVNCQFRLGQEC 540  
DB 481 PWQLFRNPHOALLHTANRPEDECVGGLACHOL-CARGHCGPGTCCVNCQFRLGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
QY 601 PQYIKANSKEFIGITELPDEBACQPCPNCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660  
DB 601 PSQVXPDLSTYMPWKFPDEBACQPCPNCTHSCVDLDDKGCAPAEQASPLTSTVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRQOKIRKTYMRLLOTELVEPLTPSCAMPNQAMRLKETEL 720  
DB 661 ILLVVVLGVVFGILIKRQOKIRKTYMRLLOTELVEPLTPSCAMPNQAMRLKETEL 720  
QY 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780  
DB 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780  
QY 781 YVSRLLGICLTSTVQLVTQMPYGCLLDHHVNRGLSGDQLLNCWQIAKGSYLESDVR 840  
DB 781 YVSRLLGICLTSTVQLVTQMPYGCLLDHHVNRGLSGDQLLNCWQIAKGSYLESDVR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRFT 900

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLKGERLPQPPICITIDVYMIWKCM 960  
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLKGERLPQPPICITIDVYMIWKCM 960  
 QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020  
 DB 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020  
 QY 1021 EYLVPQGGFFCPDPAPGAGGMVHRHRSSTRSRGGDLTLGLEPSEBAPSPAPSEG 1080  
 DB 1021 EYLVPQGGFFCPDPAPGAGGMVHRHRSSTRSRGGDLTLGLEPSEBAPSPAPSEG 1080  
 QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
 DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
 QY 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLSPKNGVWVDVFAFGAVENPEYLTPO 1200  
 DB 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLSPKNGVWVDVFAFGAVENPEYLTPO 1200  
 QY 1201 GGAAPQPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1255  
 DB 1201 GGAAPQPHPPAFSPAFDNLVYWDQDPPRGAPPSTFKGTPTAENPEYLGLDVVPV 1255

## RESULT 7

AAB21198

ID AAB21198 standard; protein; 1255 AA.

XX AAB21198;

AC AAB21198;

XX AAB21198;

DT 12-JAN-2001 (first entry)

XX Human HER-2/neu protein.

XX Human HER-2/neu protein.

XX Human HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;

XX breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX colon cancer.

XX Homo sapiens.

XX Homo sapiens.

XX WO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheyene D;

XX WPI; 2000-505976/45.

XX N-PSDB; AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of

XX the tyrosine kinase family of receptor-like glycoproteins and shows

XX homology to the epidermal growth factor receptor (EGFR). It probably

XX plays a part in cell growth and/or differentiation. The HER-2/neu

XX gene is an oncogene. An HER-2/neu fusion protein comprising a

XX HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation

XX domain may be used to treat or prevent cancer by eliciting or

XX enhancing an immune response to the HER-2/neu protein. It may be used

XX to treat malignancies such as breast, ovarian, colon, lung and

XX prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 97.8%; Score 6651; DB 21; Length 1255;  
 Best Local Similarity 97.6%; Pred. No. 0;  
 Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;  
 QY 1 MELAALCWGULLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYQCCOVQGNL 60  
 DB 1 MELAALCWGULLALLPPGAASCTGCTDMKRLPASPETHLDMRLHLYQCCOVQGNL 60  
 QY 61 ELTYLPTNASLSFLQDIOEVQGYVLLIAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLLIAHQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
 QY 121 DPLNTTPTVTCASPGGLRELQRLSUTELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
 DB 121 DPLNTTPTVTCASPGGLRELQRLSUTELKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
 QY 181 LTLIDTNRSRACHPCSPMKSGRCSGESSEDCQSLTRTVCAGGCARCKGPLTDCHEOC 240  
 DB 181 LTLIDTNRSRACHPCSPMKSGRCSGESSEDCQSLTRTVCAGGCARCKGPLTDCHEOC 240  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARCYGLGMQVYIKANSKPIGIT 360  
 DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSPCARCYGLGMQVYIKANSKPIGIT 360  
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDLSLP 420  
 DB 361 IQEFGAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDLSLP 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGLALIHNNTHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGLALIHNNTHLCFVHTV 480  
 QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQBC 540  
 DB 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCVCNCSQFLRGQBC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600  
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPPECVARC 600  
 QY 601 PQYIKANSKFIGITELPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660  
 DB 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660  
 QY 661 ILLVVVLGVTFGILIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
 DB 661 ILLVVVLGVTFGILIKRRQOKIRKYTMRLRLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
 QY 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEIIDEAIVMAGVGSPP 780  
 DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVURENTSPKANKEIIDEAIVMAGVGSPP 780  
 QY 781 YVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRCGLSQDILLNMCQIAKMSYLEVDVR 840  
 DB 781 YVSRLLGICLTSTVOLVTQLMPYGCCLLDHVRNRCGLSQDILLNMCQIAKMSYLEVDVR 840  
 QY 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRRPT 900  
 DB 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRRPT 900  
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLKGERLPQPPICITIDVYMIWKCM 960  
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREI PDLLKGERLPQPPICITIDVYMIWKCM 960  
 QY 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020  
 DB 961 IDSECRPRFRELVSFSSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA 1020



Db 961 INSECPRERELVSERARDPQRFVVIQNEIDLGPASPLDSTFYKSLLEDDMDGLVDA 1020  
QY 1021 EBYLVPQQGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSEG 1080  
Db 1021 EBYLVPQQGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDPVLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NQPDVREPQPPSPREGPLPAARPAAGATLERAKTILSPCKNGVVDVFAFGGAVENPEYLTQ 1200  
Db 1141 NQPDVREPQPPSPREGPLPAARPAAGATLERPKTILSPCKNGVVDVFAFGGAVENPEYLTQ 1200  
QY 1201 GGAAPOPHPPAFSPAFDNLTYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPOPHPPAFSPAFDNLTYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 8  
AAY84780  
ID AAY84780 standard; Protein; 1255 AA.  
XX  
AC AAY84780;  
XX  
DT 08-AUG-2000 (first entry)  
XX  
DE Amino acid sequence of the SPLICE erbb-2 receptor protein.  
XX  
KW SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KW bone resorption; inflammatory disease; degenerative disorder;  
KW wound healing.  
XX  
OS Homo sapiens.  
XX  
PN WO200020579-A1.  
XX  
PD 13-APR-2000.  
XX  
PF 01-OCT-1999; 99WO-CA00912.  
XX  
PR 02-OCT-1998; 98US-0165192.  
XX  
PA (UYMC-) UNIV MCMASTER.  
XX  
PI Muller WJ, Siegel PM;  
XX  
DR WPI; 2000-303768/26.  
XX  
Nucleic acid encoding an erbb 2 receptor protein designated SPLICE  
erbb-2, inhibitors of the protein are useful for treatment of cancer -  
Claim 3; Fig 2; 60pp; English.

The present sequence represents a SPLICE erbb-2 receptor protein. The protein has an in-frame deletion of 16 amino acids, 2 of which are conserved cysteine residues, compared to the unspliced protein. The erbb-2 polynucleotide is used to construct probes for detecting disorders of cell transformation such as cancer. Antibodies to the protein may be used to detect SPLICE erbb-2 in a sample. Agents (e.g. antisense oligonucleotides) which inhibit the expression of SPLICE erbb-2 are useful for reducing tumor cell proliferation and treating cancer. Substances which stimulate SPLICE erbb-2 are useful for treating conditions involving damaged cells including conditions in which degeneration of tissue occurs, such as arthropathy, bone resorption, inflammatory diseases, degenerative disorders of the central nervous system and wound healing.

Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 21; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;  
QY 1 MELAALCRNGLLLALLPPGAASSTOVCCTGDMKLRLLPASPEHLDMRLHLYQGCVVQGNL 60  
Db 1 MELAALCRNGLLLALLPPGAASSTOVCCTGDMKLRLLPASPEHLDMRLHLYQGCVVQGNL 60  
QY 61 ELYLPTNASLFLQDIQEVQGVYLIAHQVQVPLQRLIRIVRGTOLEFENYALAVLDNG 120  
Db 61 ELYLPTNASLFLQDIQEVQGVYLIAHQVQVPLQRLIRIVRGTOLEFENYALAVLDNG 120  
QY 121 DPLNNTTPTVGASPGGLRELOLRSLEILKGGVLIQORNPOLCYQDITLWKDILFHKNQOLA 180  
Db 121 DPLNNTTPTVGASPGGLRELOLRSLEILKGGVLIQORNPOLCYQDITLWKDILFHKNQOLA 180  
QY 181 LTLIDTNRSRACHPCSPCKGSRGCESEDQSLTRTVCAAGCARGCKPLPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPCKGSRGCESEDQSLTRTVCAAGCARGCKPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMYIKANSKFIGIT 360  
QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVFEETLEETGLYIISAWDSLP 420  
Db 361 IQEFGAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOVFEETLEETGLYIISAWDSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480  
QY 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVCNCSOFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPGPTQCVCNCSOFLRGQEC 540  
QY 541 VEECRVLOGLPREYVYNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFFCVARC 600  
Db 541 VEECRVLOGLPREYVYNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFFCVARC 600  
QY 601 PQYIKANSKFIGITELPDEEGACQPCPINCETHSCVDLDDKGPAPORASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCETHSCVDLDDKGPAPORASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMAGVQSP 780  
Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMAGVQSP 780  
QY 781 YVSRLLGICLSTVOLTPQLMPYGCLLDHRVNRGLSQDILNNCMQIAGKMSYLEDVYR 840  
Db 781 YVSRLLGICLSTVOLTPQLMPYGCLLDHRVNRGLSQDILNNCMQIAGKMSYLEDVYR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRT 900  
QY 901 HQSDVMSYGVVTWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVTYMWVKWM 960  
Db 901 HQSDVMSYGVVTWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTIDVTYMWVKWM 960  
QY 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYKSLLEDDMDGLVDA 1020  
Db 961 IDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYKSLLEDDMDGLVDA 1020  
QY 1021 EBYLVPQQGFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLPESEEAAPRSLAPSEG 1080



```
Db      1021  EYLVVPOQGFCCDPAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy      1081  AGSDVFDGDLGMAAAGLQSLPHDPSPLQRYSEDTVPULPSDTGTVAPLTCSPQPEYV 1140
Db      1081  AGSDVFDGDLGMAAAGLQSLPHDPSPLQRYSEDTVPULPSDTGTVAPLTCSPQPEYV 1140
Qy      1141  NOPDVROPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTQ 1200
Db      1141  NOPDVROPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTQ 1200
Qy      1201  GGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTTAENPEYLGLDVVP 1255
Db      1201  GGAAPQHPHPPAFSPAFDNLYYWDQPPPERGAPPSTFKGTPTTAENPEYLGLDVVP 1255
```

## RESULT 9

```
AAB85458
ID      AAB85458 standard; Protein; 1255 AA.
AC      AAB85458;
XX
XX      25-SEP-2001 (first entry)
DT
DE      Human HER-2/neu protein.
XX
XX      Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW      oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
XX      Homo sapiens.
OS
XX
XX      WO200153463-A2.
PN
XX
XX      26-JUL-2001.
XX
XX      19-JAN-2001; 2001WO-US01850.
XX
XX      21-JAN-2000; 2000US-0177545.
XX
XX      (CORI-) CORIXA CORP.
XX
XX      Cheever MA, Hand-Zimmermann S;
PI
XX
XX      WPI; 2001-476112/51.
DR      N-PSDB; AAH23392.
XX
XX      New antigen-presenting cells, useful as vaccines for eliciting or
PT      enhancing an immune response to HER-2/neu protein, particularly useful
PT      for treating or preventing cancer, e.g. breast cancer -
PS
PS      Claim 2; Page 41-46; 49pp; English.
XX
XX      The invention provides an isolated antigen-presenting cell, which
CC      expresses at least an immunogenic portion of a polypeptide that produces
CC      an immune response to HER-2/neu protein. The antigen-presenting cells are
CC      useful as vaccines for eliciting or enhancing an immune response to
CC      HER-2/neu protein, particularly in treating or preventing malignancies in
CC      which the HER-2/neu oncogene is associated. Specifically, these are
CC      useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC      colon, lung or prostate cancers. The present sequence represents
CC      the human HER-2/neu protein (also known as p185 or c-erbB2).
```

```
XX
SQ      Sequence 1255 AA;
Query Match 97.8%; Score 6651; DB 22; Length 1255;
Best Local Similarity 97.6%; Pred. No. 0;
Matches 1255; Conservative 10; Mismatches 20; Indels 0; Gaps 0;
```

```
Qy      1  MELAALCRWGLLLALLPPGAASQTQCTGTDMLRPLASPETHDMLRHLHYQGQVVGQNL 60
Db      1  MELAALCRWGLLLALLPPGAASQTQCTGTDMLRPLASPETHDMLRHLHYQGQVVGQNL 60
Qy      61  ELTYLPTNASLSFLQDIQEVQGVVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
```

```
Db      61  ELTYLPTNASLSFLQDIQEVQGVVLIHAHQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
Qy      121  DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLHKDILFKKNOLA 180
Db      121  DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIORNPOLCYQDITLHKDILFKKNOLA 180
Qy      181  LTLIDTNRSRACHPCSPCKGRSCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHQEC 240
Db      181  LTLIDTNRSRACHPCSPCKGRSCWGESSEDCOSLTRTVCAGGCARCKGPLPTDCCHQEC 240
Qy      241  AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Db      241  AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy      301  YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGQYIKANSKFITG 360
Db      301  YNYLSTDVGSCTLVCPHLHQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRVTSAN 360
Qy      361  ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLISAWPDSL 420
Db      361  IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLISAWPDSL 420
Qy      421  DLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLSLRELGLSGLALIIHNTHLCFVHT 480
Db      421  DLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLSLRELGLSGLALIIHNTHLCFVHT 480
Qy      481  PWDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNCSOFLRQEC 540
Db      481  PWDQLFRNPHOALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVCNCSOFLRQEC 540
Qy      541  VEECRVLOGLPREYVNNARHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Db      541  VEECRVLOGLPREYVNNARHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC 600
Qy      601  POYIKANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVG 660
Db      601  PSGVKPDLSPYIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVG 660
Qy      661  ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMILKETEL 720
Db      661  ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMILKETEL 720
Qy      721  RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVSP 780
Db      721  RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVSP 780
Qy      781  YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSGODLLNWCNQIAKGMYSYLEDVR 840
Db      781  YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSGODLLNWCNQIAKGMYSYLEDVR 840
Qy      841  LVHRDLAARNVLKSPNHVKITDPGLARLLIDETEHADGGKVPKMMALESILRRFT 900
Db      841  LVHRDLAARNVLKSPNHVKITDPGLARLLIDETEHADGGKVPKMMALESILRRFT 900
Qy      901  HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Db      901  HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960
Qy      961  IDSECRPRFRELVSFESRMARDQRFVITQNEIDLGSPASPLDSTFYRSLDEDDDMGLVDA 1020
Db      961  IDSECRPRFRELVSFESRMARDQRFVITQNEIDLGSPASPLDSTFYRSLDEDDDMGLVDA 1020
Qy      1021  EYLVVPOQGFCCDPAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Db      1021  EYLVVPOQGFCCDPAPAGGVMVHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
Qy      1081  AGSDVFDGDLGMAAAGLQSLPHDPSPLQRYSEDTVPULPSDTGTVAPLTCSPQPEYV 1140
Db      1081  AGSDVFDGDLGMAAAGLQSLPHDPSPLQRYSEDTVPULPSDTGTVAPLTCSPQPEYV 1140
Qy      1141  NOPDVROPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPEYLTQ 1200
```

Db 1141 NOPDVRPQPPSPREGPLPAARAGATLERPKTLSPKNGVVKDVFAGGAVENPEYLTPQ 1200  
QY 1201 GGAARQPPPPAFSAFNLVYWDQDPPRGAPPSTFKTPTAENPEYLGLDVVP 1255  
Db 1201 GGAARQPPPPAFSAFNLVYWDQDPPRGAPPSTFKTPTAENPEYLGLDVVP 1255

## RESULT 10

AAG88267  
ID AAG88267 standard; Protein; 1255 AA.

AC AAG88267;

DT 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

DE Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;  
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

OS WO200141787-A1.

PN 14-JUN-2001.

PD 11-DEC-2000; 2000WO-US33591.

PR 10-DEC-1999; 99US-0458299.

XX (EPTM-) EPIMMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
PI Keogh E;

XX WPI; 2001-374995/39.

PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
PT cellular immune responses for the prevention and treatment of cancer -  
XX Disclosure; Page 15; 199pp; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).  
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
CC culture in vitro and binds to a complex of an epitope (I), bound to a  
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
CC and a second epitope and the peptide is less than 50 contiguous amino  
CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising  
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic  
CC and immunostimulant activities, and can be used in vaccines. (I), (II)  
CC and (III) are useful for inducing cellular immune responses for the  
CC prevention and treatment of cancer. (I) and (II) are useful for  
CC monitoring or evaluating an immune response to a tumour-associated  
CC antigen when incubated with a T lymphocyte sample from a patient and  
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
CC based vaccines mean that immunosuppressive epitopes that may be present  
CC in whole antigens may be avoided. Selected epitopes may be combined to  
CC enhance immunogenicity. The possible pathological side effects caused by  
CC infectious agents or whole protein antigen is eliminated. The vaccine  
CC provides the ability to direct and focus an immune response to multiple  
CC selected antigens from the same pathogen. Epitope-based anti-tumour  
CC vaccines provides the opportunity to combine epitopes derived from  
CC multiple tumour-associated molecules addressing the problem of tumour-  
CC tumour variability and reducing the likelihood of tumour escape due to  
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
CC the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 22; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLMLRLHLYQGQVQVGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPEHLMLRLHLYQGQVQVGNL 60

QY 61 ELTYLPTNASLFLQDIQEVQVYLIAHNOVRQVPLQRLIRIVRGTOLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLFLQDIQEVQVYLIAHNOVRQVPLQRLIRIVRGTOLFEDNYALAVLDNG 120

QY 121 DPLNNTTPTVGTASPGGLREQLRSLEILKGGVLIQORNPQLCYQDILWKDIFHKNNOLA 180  
Db 121 DPLNNTTPTVGTASPGGLREQLRSLEILKGGVLIQORNPQLCYQDILWKDIFHKNNOLA 180

QY 181 LTLIDNRSRACHPCSPCKGSRGCESEDQSLTRTVCCAGCACRCKGPLPDDCCEQC 240  
Db 181 LTLIDNRSRACHPCSPCKGSRGCESEDQSLTRTVCCAGCACRCKGPLPDDCCEQC 240

QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300

QY 301 YNYLSTDVGSCTLVCPHNLNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHNLNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360

QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEOLQVFETLEETIGVLYISAWPDSL 420  
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEOLQVFETLEETIGVLYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLSLIHHNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLSLIHHNTHLCFVHTV 480

QY 481 PWDQLFRPHQALHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540  
Db 481 PWDQLFRPHQALHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540

QY 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFCVARC 600

QY 601 POYIKANSKFIGITELPDEEGACQPCINCTHSCVDLDDKGPAPAEORASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSYMFIMKFPDEEGACQPCINCTHSCVDLDDKGPAPAEORASPLTSIVSAVVG 660

QY 661 ILLVVVLGVVFGILIKRRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMAGVGP 780  
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYMAGVGP 780

QY 781 YVSRLLGLCTSTVOLVTQMLPYGCLLDHVRNRRGLSGDQLNWCQIAGKMSYLEDVDR 840  
Db 781 YVSRLLGLCTSTVOLVTQMLPYGCLLDHVRNRRGLSGDQLNWCQIAGKMSYLEDVDR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMWALESILRRRFT 900

QY 901 HQSDVMSXGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPTCTIDVYMWKVM 960  
Db 901 HQSDVMSXGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPPTCTIDVYMWKVM 960

QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
Db 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020

QY 1021 BEYLVPQGGFTCPDPAAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
Db 1021 BEYLVPQGGFTCPDPAAGAGGVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080

Db	1021	EEYLVFQQGFFCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG	1080
Qy	1081	AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ	1200
Db	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTQ	1200
Qy	1201	GGAAPQHPPPAPSPAFDNLYYWDQDPPRGAPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPPPAPSPAFDNLYYWDQDPPRGAPSTFKGTPTAENPEYLGLDVVP	1255

RESULT 11

AAE24067	AAE24067 standard; Protein; 1255 AA.
ID	AAE24067
AC	AAE24067;
XX	23-SEP-2002 (first entry)
DT	Human Her-2 protein.
DE	
XX	Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
KW	hyperproliferative disorder; prophylaxis; inflammation; antisense;
KW	tumour; gene therapy; phosphorothioate backbone.
XX	
OS	Homo sapiens.
XX	
PN	WO200222636-A1.
XX	
PD	21-MAR-2002.
XX	
PF	12-SEP-2001; 2001WO-US28572.
PR	15-SEP-2000; 2000US-0663834.
XX	
PA	(ISIS-) ISIS PHARM INC.
XX	
PI	Bennett CF, Cowsert LM;
XX	
DR	WPI; 2002-471192/50.
XX	
PT	N-PSDB; AAD38904.
PT	Novel antisense oligonucleotide which modulates the expression of Human
PT	Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT	inflammation or to prevent infection in humans -
XX	
PS	Example 13; Page 95-107; 116pp; English.
XX	
CC	The invention relates to antisense compounds targetted to a nucleic
CC	acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC	that specifically hybridises with and inhibits the expression of Her2.
CC	Antisense compounds of the invention are used for treating diseases or
CC	conditions associated with Her2 such as hyperproliferative disorders
CC	e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC	neural or cardiac cancer. They are also useful prophylactically e.g.
CC	to prevent or delay infection, inflammation and tumour formation. The
CC	invention is also used in gene therapy. The present sequence is human
CC	Her-2 protein.
XX	
SQ	Sequence 1255 AA;

Query Match	97.8%; Score 6651; DB 23; Length 1255;
Best Local Similarity	97.8%; Pred. No. 0;
Matches 1255; Conservative	10; Mismatches 20; Indels 0; Gaps 0;
Qy	1 MELAALCRWGLLIALLPFGAASQVCTGTDMLKURLPASPEHLDMLRHLYQGCVVQGNL 60
Db	i MELAALCRWGLLIALLPFGAASQVCTGTDMLKURLPASPEHLDMLRHLYQGCVVQGNL 60

Qy	61	ELTYLPTNASLSFLQDIQEQGYVLIHNOVROVPLQRLIRVGTQOLFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLSFLQDIQEQGYVLIHNOVROVPLQRLIRVGTQOLFEDNYALAVLDNG	120
Qy	121	DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA	180
Db	121	DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA	180
Qy	181	LTUIDNTRSRACHPCSPMCKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHQC	240
Db	181	LTUIDNTRSRACHPCSPMCKGSCWGESSEDCOSLRTVTCAGGCARCKGPLPTDCCHQC	240
Qy	241	AACTGCPKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AACTGCPKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMOYIKANSKFITIGIT	360
Db	301	YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN	360
Qy	361	ELFAGACKIFGSLAFLPESFDGDPASNTAPLOEQLOVEETLEEITGYLYISAWPDSLP	420
Db	361	IQEFAGCKXIFGSLAFLPESFDGDPASNTAPLOEQLOVEETLEEITGYLYISAWPDSLP	420
Qy	421	DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLSLRSLRELGLALIHNTHLFCFVHTV	480
Db	421	DLSVFQNLQVIRGRILHNGAYSLTLOGLGISWGLSLRSLRELGLALIHNTHLFCFVHTV	480
Qy	481	PWDQLFRNHQALLHTANRPEDECVGEGLAHQLCARGHCWGPCTQVNCQSLRQOEC	540
Db	481	PWDQLFRNHQALLHTANRPEDECVGEGLAHQLCARGHCWGPCTQVNCQSLRQOEC	540
Qy	541	VEECRVLOGLPREYVNHARHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC	600
Db	541	VEECRVLOGLPREYVNHARHCLPCHPECQONGSVTCFGEADOCVACAHYKDPFPCVARC	600
Qy	601	POYIKANSKFITIGITELPDEEGACQPCINCTHSCVDLDDKCPAEORASPLTSTVSAVVG	660
Db	601	PSGKPDLSYMPIWKFDEEGACQPCINCTHSCVDLDDKCPAEQASPLTSTISAVVG	660
Qy	661	ILLVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRILKETEL	720
Db	661	ILLVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQOMRILKETEL	720
Qy	721	RKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP	780
Db	721	RKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVSP	780
Qy	781	YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGODLLNWCQIAKGMYSYLEDVR	840
Db	781	YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGODLLNWCQIAKGMYSYLEDVR	840
Qy	841	LVRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT	900
Db	841	LVRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRFT	900
Qy	901	HOSDVMWSYGVVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCM	960
Db	901	HOSDVMWSYGVVWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCM	960
Qy	961	IDSECRPRFRELVSFSEARMARDPQREVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA	1020
Db	961	IDSECRPRFRELVSFSEARMARDPQREVIQNEDLGPASPLDSTFYRSLLDEDDMDGLVDA	1020
Qy	1021	EYLVVPOQGFPCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG	1080
Db	1021	EYLVVPOQGFPCDDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG	1080
Qy	1081	AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ	1200

Db 1141 NQPDVRQPSREGPLPAARAGATLERPKTSLSGKNGVVKDVFARGGAVENPEYITPQ 1200  
Qy 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255  
RESULT 12  
AAE20479  
ID AAE20479 standard; Protein; 1255 AA.  
AC AAE20479;  
XX  
DT 01-JUL-2002 (first entry)  
XX Human Her-2/neu protein.  
DE  
XX  
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
KW Human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 1021..1030  
FT /note= "Naturally processed HLA-B44-restricted epitope"  
XX  
XX WO200214503-A2.  
XX 21-FEB-2002.  
XX  
XX 14-AUG-2001; 2001WO-US41733.  
XX  
XX 14-AUG-2000; 2000US-225152P.  
PR 28-SEP-2000; 2000US-236428P.  
XX 21-FEB-2001; 2001US-270520P.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Hand-zimmermann S, Cheever MA, Foy TW, Lodes MJ, Kalos MD;  
PI McNeill PD, Vedvick TS;  
XX  
XX WPI; 2002-280758/32.  
DR N-PSDB; AAD32743.  
XX  
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
PT prevention and diagnosis of cancer, preferably breast cancer  
XX  
XX Disclosure; Page 114-117; 129pp; English.  
XX  
XX The invention relates to an isolated Her-2/Neu polypeptide composition  
CC effective for eliciting an immune response. The invention is useful for  
CC eliciting an immune response in a patient, where the patient is human  
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
CC The composition is useful for the therapy and diagnosis of cancer,  
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
CC and other compositions for the diagnosis, prevention and treatment of  
CC human malignancies, for stimulating and/or expanding T cells specific for  
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
CC patient. The invention is useful for stimulating a T cell response in a  
CC human patient, as probe or primer for nucleic acid hybridisation to  
CC selectively form duplex molecules with complementary stretches of the  
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
CC length gene from a suitable library, and to direct expression of a  
CC polypeptide in appropriate host cells. The composition is useful in  
CC prophylactic or therapeutic applications and for the treatment of cancer,  
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
CC associated malignancies. The invention is useful in gene therapy. The  
CC present sequence is human Her-2/neu protein.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.8%; Score 6651; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLLALLPAGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60  
Db 1 MELAALCRWGLLLALLPAGAASTQVCTGDMKRLPASPETHLDMRLHLYGQCQVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIOEVQGVYVLI AHNQVRQVPLQRLRIVRGQTQLPEDNYALVDNG 120  
Db 61 ELTYLPTNASLSFLQDIOEVQGVYVLI AHNQVRQVPLQRLRIVRGQTQLPEDNYALVDNG 120  
Qy 121 DPLNNTTPTVGASPGGLRELQRLSITLILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180  
Db 121 DPLNNTTPTVGASPGGLRELQRLSITLILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGSESSDCQSLTRTVACAGGACRCKGPLPTDCCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKGSRGSESSDCQSLTRTVACAGGACRCKGPLPTDCCCHQC 240  
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGHQYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGHQYIKANSKFIGIT 360  
Qy 361 ELEFAGCKKIFGSLAFPLESFDGDPASNTAPLOEQLOVFTLEETITGLYVISAWPDSLP 420  
Db 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPLOEQLOVFTLEETITGLYVISAWPDSLP 420  
Qy 421 DLSVFQNLQVIRGIRLHNGAYSLSLTLQGLIGISWGLRSLRELGLGSLALIHNTLFCVHTV 480  
Db 421 DLSVFQNLQVIRGIRLHNGAYSLSLTLQGLIGISWGLRSLRELGLGSLALIHNTLFCVHTV 480  
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCHGPGPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCHGPGPTQCVNCSQFLRGQEC 540  
Qy 541 VEBCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600  
Db 541 VEBCRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAVRC 600  
Qy 601 POYIKANSKFIGITELPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVVG 660  
Db 601 PSQVKPDLSPYIWKPPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTISVAVVG 660  
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVSP 780  
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVMAGVSP 780  
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840  
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGQDLNWCMIKAGMSYLEDVR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900  
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDYMIWKVCM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDYMIWKVCM 960  
Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVA 1020  
Qy 1021 EYLVVQQOFFFCDDPAPGAGGMVHRHSSTSGGDLTLGLEPSEEEAPRPLAPSEG 1080  
Db 1021 EYLVVQQOFFFCDDPAPGAGGMVHRHSSTSGGDLTLGLEPSEEEAPRPLAPSEG 1080

Db 1021 EYLVPOQGFCDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGLGMAAGKQSLPDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGLGMAAGKQSLPDPSPLORYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPQ 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLTPQ 1200  
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

## RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

AC

XX AAM51143;

DT 17-JUN-2002 (first entry)

XX

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX

KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain

FT 1..653

FT /note= "extracellular domain"

FT Domain

FT 676..1255

FT /note= "intracellular domain"

FT Domain

FT 990..1255

FT /note= "phosphorylation domain"

XX

PN WO200212341-A2.

XX

XX 14-FEB-2002.

XX

XX 03-AUG-2001; 2001WO-US24283.

XX

XX 03-AUG-2000; 2000US-0632507.

XX

XX (CORI-) CORIXA CORP.

XX

XX (SMK) SMITHKLINE BEECHAM BIOLOGICALS.

XX

XX Cheever MA, Gheysen D;

XX

XX WPI; 2002-241743/29.

XX

XX N-PSDB; ABA92250.

XX

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX

XX or enhancing an immune response to the protein, has Her-2/neu

XX

XX phosphorylation domain

XX

XX Claim 68; Fig 7; 141pp; English.

XX

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX

CC comprising the fusion proteins or nucleic acid molecules. In  
CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
CC protein is fused to a Her-2/neu intracellular domain or  
CC phosphorylation domain (or its DeltaPp fragment). An immune  
CC response to Her-2/neu protein is elicited or enhanced by  
CC administering the fusion protein in the form of a vaccine, or by  
CC transfecting cells of an animal ex vivo with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

SQ

Query Match 97.8%; Score 6651; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHYQCQVVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPETHLDMRLHYQCQVVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVROVPLQRLRVRGQTQLPEDNYVALVDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVROVPLQRLRVRGQTQLPEDNYVALVDNG 120

Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGGVLTQPNQCYQDTILWKDILFHNKQILA 180

Db 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGGVLTQPNQCYQDTILWKDILFHNKQILA 180

Qy 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARGPLPTDCHEQC 240

Db 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTVCAAGCARGPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360

Db 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIYISAMPDSL 420

Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIYISAMPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLSRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVCNCSQFLRGQBC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPGPTQCVCNCSQFLRGQBC 540

Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Qy 601 PQYIKANSKFITIGITELPDEEGACQPCPINCNTHTSCVDLDDKGCAPORASPLTISAVVG 660

Db 601 PSQVXPDLSTYMPIWKFPDEEGACQPCPINCNTHTSCVDLDDKGCAPORASPLTISAVVG 660

Qy 661 ILLVVLGVWFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVLGVWFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

Db 721 RKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRGRGLSGDOLLNWCQIAKGSYLEDVD 840  
DB 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRGRGLSGDOLLNWCQIAKGSYLEDVD 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRFT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRFT 900  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960  
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960  
QY 961 IDSECRPRFRELSEFSRMDRPFVVIQNEIDLGRASPLDSTFYRSLLEDGMDGLVDA 1020  
DB 961 IDSECRPRFRELSEFSRMDRPFVVIQNEIDLGRASPLDSTFYRSLLEDGMDGLVDA 1020  
QY 1021 EYLVPOQGFCCPDPAFGAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLAPSEG 1080  
DB 1021 EYLVPOQGFCCPDPAFGAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRSPPLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVYAPLTCSPQPEYV 1140  
QY 1141 NQPDVVRPQPSPREGLPLAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200  
DB 1141 NQPDVVRPQPSPREGLPLAARPAAGATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPAPFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255  
DB 1201 GGAAPQHPHPAPFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

## RESULT 14

AAU77114

ID AAU77114 standard; Protein; 1255 AA.

AC AAU77114;

XX 05-JUN-2002 (first entry)

DT Human Her-2/neu polypeptide.

XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;

XX acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;

XX chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;

XX Hodgkin's lymphoma; T cell therapy.

XX Homo sapiens.

OS WO200213847-A2.

XX 21-FEB-2002.

PD 13-AUG-2001; 2001WO-US25408.

XX 14-AUG-2000; 2000US-0638280.

XX 28-SEP-2000; 2000US-0675904.

XX (CORI-) CORIXA CORP.

PA Gaiger A, Cheever MA, Hand-zimmermann S;

XX WPI; 2002-280741/32.

XX N-PSDB; ABK10730.

XX Inhibiting haematological malignancy development by administering

PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide

PT encoding the polypeptide, or antigen presenting cells expressing the

PT polypeptide

XX Disclosure; Page 71-74; 74pp; English.

PS

XX

CC The invention relates to a method for inhibiting development of

CC haematological malignancy in a patient by administering a polypeptide

CC comprising an immunogenic portion of Her-2/neu or a polynucleotide

CC encoding the polypeptide. Antigen presenting cells that express the

CC protein can also be administered. The sequences are used for inhibiting

CC development of haematological malignancy such as acute myelogenous

CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic

CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's

CC lymphoma. This sequence represents the human Her-2/neu polypeptide.

XX

SQ Sequence 1255 AA;

Query Match 97.8%; Score 6651; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVVQGNL 60

DB 1 MELAALCRWGLLALLPPGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVVQGNL 60

QY 61 ELTYLPTNASLSFLQDIQEYGVYLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASLSFLQDIQEYGVYLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180

DB 121 DPLNNTPTVTGASPGGLRELQSLTEILKGGVLIQRNPOLCYQDTTLWKDIFHNQOLA 180

QY 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240

DB 181 LTLIDNRSRACHPCSPMKSGSCWGESSEDCOSLTRTVCAGGCARCKGLPTDCCHEQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGCTLVCPLNHOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFIGIT 360

DB 301 YNYLSTDVSGCTLVCPLNHOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFIGIT 360

QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVETLEEITGYLIYISAWPSLP 420

DB 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLOEQLQVETLEEITGYLIYISAWPSLP 420

QY 421 DLSVFQNLQVIRGRIILHNGAYSILTLOGLISWLSRLSRLGSLALIHNTLHLCFVHTV 480

DB 421 DLSVFQNLQVIRGRIILHNGAYSILTLOGLISWLSRLSRLGSLALIHNTLHLCFVHTV 480

QY 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540

DB 481 PWDQLFRNPHQALLHTANRPEDECEVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540

QY 541 VEBCRVLQGLPREYVNRHCLPCHPECPONGSVTCFGEADOCVACAHYKDPBFCVARC 600

DB 541 VEBCRVLQGLPREYVNRHCLPCHPECPONGSVTCFGEADOCVACAHYKDPBFCVARC 600

QY 601 POYIKANSKFIGITELPDEBEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTISAVVG 660

DB 601 PSQVXPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTISAVVG 660

QY 661 ILLVWVVGWVFGILIKRROOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720

DB 661 ILLVWVVGWVFGILIKRROOKIRKYTMRLRLQETELVEPLTPSGAMPNQAMRILKETEL 720

QY 721 RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780

DB 721 RKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRGRGLSGDOLLNWCQIAKGSYLEDVD 840

DB 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRENRGRGLSGDOLLNWCQIAKGSYLEDVD 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWKMALESILRRRFT 900

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPICKMALESILRRFT 900  
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960  
Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020  
Qy 1021 EYLVPOQGFCCPDPAFCAGMVMHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFCCPDPAFCAGMVMHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDTPVLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDTPVLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYILTPQ 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKOVFAFGGAVENPEYILTPQ 1200  
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYGLDVPV 1255

RESULT 15

AAR39568  
ID AAR39568 standard; Protein; 1433 AA.  
XX AAR39568;  
DT 07-FEB-1994 (first entry)  
XX Sequence of c-erbB-2 tumour antigen.  
XX Tumour antigen; c-erbB-2; glycoprotein.  
XX Homo sapiens.  
XX W09316185-A.  
XX 19-AUG-1993. 93WO-US01055.  
XX 05-FEB-1993; 93WO-US01055.  
XX 06-FEB-1992; 92US-0831967.  
XX (CETU ) CETUS ONCOLOGY CORP.  
XX (CREA-) CREATIVE BIOMOLECULES INC.  
XX Houston LL, Huston JS, Oppermann H, Ring DB;  
XX WPI; 1993-272889/34.  
XX DR N-PSDB; AAQ46083.  
XX New single chain Fv polypeptide binding to C-erbB-2 tumour  
XX antigen - for imaging or treating breast or ovarian cancer etc.  
XX Disclosure; pages 48-54; 87pp; English.  
XX c-erbB-2 refers to a protein antigen expressed on the surface of  
XX tumour cells, such as breast and ovarian tumour cells, which is an  
XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents  
XX the location of a stop codon in AAQ46083.  
XX Sequence 1433 AA;

Query Match 97.1%; Score 6608; DB 14; Length 1433;  
Best Local Similarity 97.1%; Pred. No. 0;

Matches 1218; Conservative 12; Mismatches 25; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVQGCVOVQGNL 60  
Db 1 MELAALCRWGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLVQGCVOVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNVQVROPVLPQRLRIVRGTLQFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIOEVQGVYLIAHNVQVROPVLPQRLRIVRGTLQFEDNYALAVLDNG 120  
Qy 121 DPLNNTTPTVGASPGGLRELQLSLTBIILKGGVLIQRPOLCYODTILWKDIFHNQOLA 180  
Db 121 DPLNNTTPTVGASPGGLRELQLSLTBIILKGGVLIQRPOLCYODTILWKDIFHNQOLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKCKSGRCWGSSEDCOSLTRTVCGAGCARCKGPLPTCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKCKSGRCWGSSEDCOSLTRTVCGAGCARCKGPLPTCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSTLVCPLHNOEVTAEADGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360  
Db 301 YNYLSTDVGSTLVCPLHNOEVTAEADGTQRCCKSPCARVCYGLGMOYIKANSKFTGIT 360  
Qy 361 ELSPAGCKKIFGSLAFLPESFDGPASNTAPLOEQLOVRETLEITGYLYISAMPDLSL 420  
Db 361 IQBFAGCKRIFGSLAFLPESFDGPASNTAPLOEQLOVRETLEITGYLYISAMPDLSL 420  
Qy 421 DLSVFQNLQVIRGILHNGAYSLTLQGLIGISWLGSLRLRELGLALIHHTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGILHNGAYSLTLQGLIGISWLGSLRLRELGLALIHHTHLCFVHTV 480  
Qy 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQFLRGQEC 540  
Db 481 PWDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQFLRGQEC 540  
Qy 541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSVTCFPGPADQCVACAHYKDPFCVARC 600  
Db 541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSVTCFPGPADQCVACAHYKDPFCVARC 600  
Qy 601 PQYIKANSKFITELPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISVAVG 660  
Db 601 PSGVKPDLSTYMPIWKIPDEEGACQPCPINCTHSCVDLDDKGCPAEQRASPLTISVAVG 660  
Qy 661 ILLVVLGVVFGILIKRROQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVLGVVFGILIKRROQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMVAGVSP 780  
Db 721 RKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMVAGVSP 780  
Qy 781 YVSRLLGICLTSTVQLVTQLMYPYGLLDHVRENRRGLSQDLLNWCMIKAGMSYLEDVR 840  
Db 781 YVSRLLGICLTSTVQLVTQLMYPYGLLDHVRENRRGLSQDLLNWCMIKAGMSYLEDVR 840  
Qy 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPICKMALESILRRFT 900  
Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGGKVPICKMALESILRRFT 900  
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVKCWM 960  
Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020  
Qy 1021 EYLVPOQGFCCPDPAFCAGMVMHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFCCPDPAFCAGMVMHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVVRQPPSPREGPLPAARAGATLERAKTILSPGKNGVWKDVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVVRQPPSPREGPLPAARAGATLERAKTILSPGKNGVWKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:41:45  
 Job time : 42.9774 secs



GenCore version 5.1.6

Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds

(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-579-593-12

Perfect score: 6803

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVFPV 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mbc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6147	90.4	1259	6	O18735 canis famil
2	3112	45.7	1209	11	Q9X70 rattus norv
3	3087	45.4	1210	11	Q9EP98 mus musculu
4	2698	39.7	1165	13	Q9YH40 xiphophorus
5	2680.5	39.4	1137	13	Q9W6F6 gallus gall
6	2280	33.5	1328	13	P79754 fugu rubrip
7	2028.5	29.8	1433	5	Q9BHH9 anopheles g
8	1871	27.5	419	4	Q9UK79 homo sapien
9	1739	25.6	367	11	Q8R2X1 mus musculu
10	1697.5	25.0	412	4	Q8WYV0 homo sapien
11	1690	24.8	729	15	Q86712 avian rous-
12	1688	24.8	567	15	Q86714 avian rous-
13	1623.5	23.9	962	15	Q64895 avian eryth
14	1615	23.7	545	15	Q85468 avian eryth
15	1478.5	21.7	655	11	Q9WVF5 mus musculu
16	1462.5	21.5	643	11	Q9ERV6 mus musculu

Q9Y1X8 ephydatia f  
Q23821 caenorhabdi  
Q26566 schistosoma  
Q90836 gallus gall  
Q98802 rattus norv  
Q98802 rattus norv  
Q14256 homo sapien  
Q923V5 rattus norv  
P11776 xiphophorus  
Q8SZW1 drosophila  
Q91622 xiphophorus  
Q9BUD7 homo sapien  
Q9PVZ4 xenopus lae  
Q9NJV5 biophalari  
Q9B666 oryctolagus  
Q93457 scophthalmu  
Q8UW85 paralicthty  
Q8UW86 paralicthty  
Q9U5A8 bombyx mori  
Q73798 xenopus lae  
Q8UW84 paralicthty  
Q8UW83 paralicthty  
Q9YGH8 scophthalmu  
Q9V894 drosophila  
Q9GVW4 rattus sp.  
Q9LYM0 mus musculu  
Q96135 homo sapien  
Q98MR2 mus musculu  
Q07912 homo sapien

## ALIGNMENTS

RESULT 1

O18735 PRELIMINARY; PRT; 1259 AA.  
AC O18735;  
DT 01-JAN-1998 (TremBLrel. 05, Created)  
DT 01-JAN-1998 (TremBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TremBLrel. 21, Last annotation update)  
DE ErB-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erbB-2 from canine mammary gland";  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1;  
DR HSSP; P11362; 1FGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transference; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4CD46 CRC64;

Query Match	90.4%;	Score 6147;	DB 6;	Length 1259;		
Best Local Similarity	90.0%;	Pred. No. 0;				
Matches 1134;	Conservative 46;	Mismatches 74;	Indels 6;	Gaps 2;		
QY	1	MELAAACRWGLLLALLPPGAAS	TQVCTGDMKRLRLPASBETHLDMRLHLYQGCQVQGNL	60		
Db	1	MELAAACRWGLLLALLPSGAAG	TQVCTGDMKRLRLPASBETHLDMRLHLYQGCQVQGNL	60		
QY	61	ELTYLPTNASLFLQDIOEVOGYVLI	AHNOVROVPLQRLRIRVGTQOLFEDNYALAVLDNG	120		
Db	61	ELTYLPANASLFLQDIOEVOGYVLI	AHSQVRQIPQLQRLRIRVGTQOLFEDNYALAVLDNG	120		
QY	121	DPLNNTTPTVGTASPGGLRELQLRSL	TEILKGGVLIQRNPOLCYQDITLWKDIFHKNQLA	180		
Db	121	DPLEGGIPAGAAQAGLRELQLRSL	TEILKGGVLIQRSPQLCHQDITLWKDVFHKNQLA	180		
QY	181	LTLIDNRSRACHPCSPMKGR	SCWGESSEDCOSLRTTVCAGGCARCKGPLPTDCCHEQC	240		
Db	181	LTLIDNRSFACPPCSACKDA	HCWGAASSGDCOSLRTTVCAGGCARCKGPQPTDCCHEQC	240		
QY	241	AAGCTGPKHSDCLACIHNHSGICEL	HCPALVTYNTDTFESMNPPEGRYTFGASCVTACP	300		
Db	241	AAGCTGPKHSDCLACIHNHSGICEL	HCPALVTYNTDTFESMNPPEGRYTFGASCVTSCP	300		
QY	301	YNYLSTDVGSCTLVCLPHNQEVTA	EDGTQRCCKSPCARVCYGLGMQYIKANSKFTGIT	360		
Db	301	YNYLSTDVGSCTLVCLPLNNQEVTA	EDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN	360		
QY	361	ELFACGKIKFGLSLAFLPESFDGP	ASNTAPLOEQLOVFTLEETGYLIYSAMDPSLP	420		
Db	361	IOEFACGKIKFGLSLAFLPESFDGP	ASNTAPLOEQLRVFEALEETGYLIYSAMDPSLP	420		
QY	421	DLISVFONLQVIRGRILNHGAYSLT	LOGLGSWGLSRLSRELGSGLAIHNHNLFCFVHTV	480		
Db	421	NLSVFONLQVIRGRVLDHGA	YSLTLOGLGSWGLSRLSRELGSGLAIHNHNLFCFVHTV	480		
QY	481	PMDQLFRNHQALLTANRPEDEVC	GEGLACHOLCARGHCGPPTQCVNCSQFLRQEC	540		
Db	481	PMDQLFRNHQALLSANRPEBEC	VEGEGLACYP-CAHGHCMWGPPTQCVNCSQFLRQEC	539		
QY	541	VEECRVLOGLPREYVNA	RHCLPCHPECPQONGSVTCFGEADQCVACAHYKDPFCVARC	600		
Db	540	VEECRVLOGLPREYVND	RYLPCHECQPQNGSVTCFGEADQCVACAHYKDPFCVARC	599		
QY	601	POVIKANSFIGITELPDEBEGAC	QCPINCTHSCVDLDDKCPAERASPLTISVSVAVG	660		
Db	600	PSGVKPDLSFMTLWKEDEBEGT	QCPINCTHSCADLDEKCPAERASPLTISIAAVG	659		
QY	661	ILLVVLGVVFGILIKRQOKIRKY	TMRLLOETELVEPLTPSGAMPNQAMRILKETEL	720		
Db	660	ILLAVVVLVLGILIKRROKIRKY	TMRLLOETELVEPLTPSGAMPNQAMRILKETEL	719		
QY	721	RKVVLGSGAGFYVYKGIWIPD	GENVKIPVAIKVLRNTSPKANKILDEAYVWAGVSP	780		
Db	720	RKVVLGSGAGFYVYKGIWIPD	GENVKIPVAIKVLRNTSPKANKILDEAYVWAGVSP	779		
QY	781	YVSRLLGICLTSTVQLVTQML	PMYGCILLDHVRENRLSGLSQDOLLNWCQIAKGSYLEDVDR	840		
Db	780	YVSRLLGICLTSTVQLVTQML	PMYGCILLDHVREHRLSGLSQDOLLNWCQIAKGSYLEDVDR	839		
QY	841	LVRDLAARNVLVKS	PNHVKITDIFGLARLLDIDETEHADGGVKVP	IKMALESILRRPT	900	
Db	840	LVRDLAARNVLVKS	PNHVKITDIFGLARLLDIDETEHADGGVKVP	IKMALESILRRPT	899	
QY	901	HQSDVMSYGVYVWELMTFGA	KPYDGIIPAREIPDLEKGERLPPOPICTIDVYMTWVKWM	960,		
Db	900	HQSDVMSYGVYVWELMTFGA	KPYDGIIPAREIPDLEKGERLPPOPICTIDVYMTWVKWM	959		
QY	961	IDSECRPRFRELVS	EFSESRMARDPQRFVVIQNE	DLGASPFLDSTFYRSLLEDDDDMGDLVDA	1020	
Db	960	IDSECRPRFRELVA	EFSESRMARDPQRFVVIQNE	DLGASPFLDSTFYRSLLEDDDDMGDLVDA	1019	
QY	1021	EYLVLPQQGFCF	PDPAFAGAGMWHRRSSSTRSGGDLT	LTGLEPSEEEAPRSLAPSEG	1080	
Db	1020	EYLVLPQQGFCF	PDPAFAGAGTAHRRSSSTRSGGDLT	LTGLEPSEEEPPKSLAPSEG	1079	
QY	1081	AGSDVFDGDLGMAAKGLQSLP	THDPSPQLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140		
Db	1080	AGSDVFDGDLGMAAKGLQSLP	SDPSPQLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1139		
QY	1141	NQPDVVRPQPPSRGGLPAAR	PAGATLER-----AKTSLSPQKNGVVKDVFAGGAVENPE	1195		
Db	1140	NQPEYVWPQPLALEGLP	PPSRPAGATLERPKTSLPKTSLSPKNGVVKDVFAGGAVENPE	1199		
QY	1196	YLTPOGGAAPPHPPPAFSP	ADNLYYWDQDPPRGAPPSFTKCTPTAENPEYLGLDVVP	1255		
Db	1200	YLAPRGAAPPHPPPAFSP	ADNLYYWDQDPSRGSPSTFEGTPTAENPEYLGLDVVP	1259		
RESULT 2						
Q9QX70	PRELIMINARY;	PRT;	1209	AA.		
AC	Q9QX70;					
DT	01-MAY-2000 (TREMBlrel. 13, Created)					
DT	01-MAY-2000 (TREMBlrel. 13, Last sequence update)					
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)					
DE	Epidermal growth factor receptor.					
GN	EGFR.					
OS	Rattus norvegicus (Rat).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
OX	NCBI_TaxID=10116;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	MEDLINE=90258888; PubMed=2342466;					
RA	Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,					
RA	Earp H.S.;					
RT	"A truncated, secreted form of the epidermal growth factor receptor is					
RT	encoded by an alternatively spliced transcript in normal rat tissue.";					
RL	Mol. Cell. Biol. 10:2973-2982(1990).					
RN	[2]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	Petch L.A.;					
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.					
RN	[3]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=FISHER; TISSUE=LIVER;					
RA	Guttridge K., Dawson T.L., Earp H.S.;					
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.					
DR	EMBL; M37394; AAF14008.1;					
DR	HSSP; P11362; 1FGK.					
DR	InterPro; IPR000494; EGFR_L domain.					
DR	InterPro; IPR000179; Euk_pkinase.					
DR	InterPro; IPR002174; Furin-like.					
DR	InterPro; IPR001245; Tyr_pkinase.					
DR	Pfam; PF00757; Furin-like; 1.					
DR	Pfam; PF00069; pkinase; 1.					
DR	Pfam; PF01030; Recep_L domain; 2.					
DR	PRINTS; PR00109; TYRKINASE.					
DR	ProDom; PD000001; Euk_pkinase; 1.					
DR	SMART; SM00261; FU; 3.					
DR	SMART; SM00219; TyrKc; 1.					
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.					
DR	PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.					
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.					
KN	ATP-binding; Receptor; Tyrosine-protein kinase.					
SQ	SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;					
Query Match	45.7%;	Score 3112;	DB 11;	Length 1209;		
Best Local Similarity	50.0%;	Pred. No. 3e-226;				
Matches 639;	Conservative 164;	Mismatches 359;	Indels 116;	Gaps 28;		
QY	3	LAALCRWGLLLALLP	PGA-ASTQVCTGTDMLRLPASBETHLDMRLHLYQGCQVQGNLE	61		

Db 15 LAALCAAG-----GALBEKKVCOGTSNRLTQLTGTFDEHFLSLQRMFNNECVVLGNLE 66  
Qy 62 LTYLPTNASLFLQDIOEQVGYVLIHANQVQVPLQRLRIVRGTLQEDYEDYALAVLNDG 121  
Db 67 ITVVQRNDLSFKITQEVAGYVLIATNVERIPLENLQIRGNALYENTYALAVLSN-- 124  
Qy 122 PLNNTTPVTGASPGGLRELQRLSRLTEILKGVLQIRNPOLCYODTILWKLDFHKNQAL 181  
Db 125 -----YGTNKTGLRELPMRLQELIIGAVFNSNPILCNMETIQMDIV-QDVFLSN 175  
Qy 182 TLIDTNRS-RACHPCSPMKGSCWGESSEDCQSLTRTVCAAGCA-RCKGLPLTDCCHEQ 239  
Db 176 MSMDVQRHLTGCPKCPDPCPNCGCWGRGEBENCKLTKIICAQCSCRCRGRSPSDCCNQ 235  
Qy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299  
Db 236 CAAGCTGPRSDCLVCHRFDENATCKDTCPLMLNPTTYQMDVNPBGKYSFGATCVKCK 295  
Qy 300 PYNLYLTDVSGSTLVCPLHMQEVTAEQTCRCKSKPCARVCYGLCMQVYKANSKFIGI 359  
Db 296 PRNVVYTDHGSVCVRACGPDYEV-EEDGVSKCKKCDGPKVCNGVIGIGEFK-DTLSINA 353  
Qy 360 TELE-FAGCKKIFGSLAFSPESFGDPASNTAPLOPEQLQVFTLEBEITGLYISAMPDS 418  
Db 354 TNIKFKYCTAISGDLHLPLVAFKGSFTRTPPLDPRELEILKTVKEITGFLIQAOPEN 413  
Qy 419 LPDLSVFQNLQVIRGLHNGAYSLTLOGLISWGLRLSRLRELGLALHHTHLCFVH 478  
Db 414 WTDLHAFENLEIIRGRTKHQGSLSLAVGLNITSLGLSLKESDGDVITSGNRNLICAN 473  
Qy 479 TVPWDLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCGPGTQCVNCSQFLRGQ 538  
Db 474 TINWKLFTPNQTKIMNRAEKDKATNHVCNPLCSSEGCWGPEDTCVSCQNVSRGR 533  
Qy 539 ECVEECRVLOGLPREVYNARHCLPCHPEQOPQNGSVTCFGEADQCVACAHYKDPFCVA 598  
Db 534 ECVDKCNILEGEPRFVENSECTQCHPECLPOTMNTICTGRGPDNCIKCAHYVDGPHCVK 593  
Qy 599 RCPQYIKANSKFIGITELPEDEGACQPCPINCTHSCVDLDDKCPAEQASRP-LTISVA 657  
Db 594 TPCSGIMGENTLV-WKFDANNVCHLCHANCTYGCAGPLKGC--QOPEGPKIPSIATG 650  
Qy 658 VVGILLVVLGVVFGI-LIKRROQKIRKVTMRLLQETELVEPLTPSGAMPNOAMRILK 716  
Db 651 IVGGLLFIVV-VALGIGLFWRRQLVRKTLRLQLRELVEPLTPSGAPNAQHURILK 709  
Qy 717 ETELKRVKVLGSAFTGYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAG 776  
Db 710 ETEFKIKVLGSAFTGYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVMAS 769  
Qy 777 VGSPPVSRLLGICLTSTVOLVTQLMBYGCLLDHVRENRLGSLQDILLNMCQIAKMSYL 836  
Db 770 VDNPHVCRLLGICLTSTVOLITQLMYPYGLLDYVREHKONIGSQYLLNMCVQIAKGMNYL 829  
Qy 837 EDVRLVHRLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILR 896  
Db 830 EDRRLVHRLAARNLVKTPQHVKITDFGLAKLLGAEKEYRAEGKVPKIMMALESILH 889  
Qy 897 RRTHQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPICTIDVYIMV 956  
Db 890 RIYHQSDVMSYGVTVWELMTGSKPYDGIPIASEISSILEKGERLPQPPICTIDVYIMV 949  
Qy 957 KCMWIDSECRPRFRELVSFBSRWARDPQRFVVIQ-NEDLGPASPLDSTYRSILLEDDMG 1015  
Db 950 KCMWIDASRPRFRELILFBSKWARDPQRYLVIQGDERMHLFSPDTSNFRALMEEDME 1009  
Qy 1016 DLVDAEYLVPOQGFPCPDPAAGAGMVHRRSSRSGGDLTLGLSPSEEAAPRSL 1075  
Db 1010 DVVDAEYLVPOGFF-----NSPST-----SRPL 1035  
Qy 1076 APSEGAGSDVFDGLGMGAAGKQLSLPHDPSPLOQRYSEDPTVPLPSET--DGYVAPLTC 1133  
Db 1036 LSSLSANSN-----SSTVACINRNGSCRVKDEAFLQRYSSDPTSVLTEDNIDDTFL----- 1086

Qy 1134 SPOPEYVNOQDVVRPQPPSPREGPLPAARAPAGATLERAKTILSPGKNGVVVKVFAFGGAVEN 1193  
Db 1087 -PVEYLINO-SVPRKPAQSVQNPVYHNOPLHP-----APGRDLHYON--PHSNVSN 1134  
Qy 1194 PEYL-TPOGAAQPPHPPAFSPAFONLYYWDQ-----DP-----PERGAPPST 1236  
Db 1135 PEYLNTAQ-----PTCLSSGPFSSALWIOKSHOMSLDNDPDYODFFPKBAKPNGI 1185  
Qy 1237 FKGTPTAENPEYLGLDVP 1254  
Db 1186 FKG-PTAENAEYLRVAPP 1202  
RESULT 3  
Q9EP98  
ID Q9EP98 PRELIMINARY; PRT; 1210 AA.  
AC Q9EP98;  
DT 01-MAR-2001 (T-EMBLrel. 16, Created)  
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Epidermal growth factor receptor isoform 1.  
EGFR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;  
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
RA Maible N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and  
RT mouse alternative Egfr transcripts encoding truncated receptor  
RT isoforms";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J;  
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
RA Maible N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and  
RT mouse alternative Egfr transcripts encoding truncated receptor  
RT isoforms";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF275366; AAG28045.1; -  
DR EMBL; AF275366; AAG28045.1; JOINED.  
DR EMBL; AF275365; AAG28045.1; JOINED.  
DR EMBL; AF275367; AAG24386.1; -  
DR HSSP; P11362; 1FGK.  
DR MGD; MGI:95294; Egfr.  
DR InterPro; IPR000345; CytC\_heme\_bind.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00220; S\_TKC; 1.  
DR SMART; SM00219; TyRK; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C\_UNKNOW\_N\_1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.

Db 956 DADSRKFKRELILLESFEMARDPORYLIVIGQDERMHLPSPTDTSFYFALMDEEDMEDVDA 1010  
 Qy 1021 EYLVPVQQGFFCPDPAPGAGGMVHRRHSSTSSGGDLTLGLEPSEEEAPRSPAPSE 1080  
 Db 1016 DEYLTTPQGGFF-----NSPST-----SRTPLLSSLS 1041  
 Qy 1081 AGSDVFDGDLGMAAGKLOSLPTHDSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQE 1138  
 Db 1042 ATSN-----NSTVACINRNGSCRYKEDAFQRYSDPTGAVTEDNIDDAFL-----VPVE 1091  
 Qy 1139 YVNPQDVRVOPPPSPREGPLPAARPAGATLERAKTLPQKGVVVDVFAFGGAVENPEYL- 1197  
 Db 1092 YVNQ-SVPRKPAGSVQNPVTHNQPLHP-----APGRDLHYQN--PHSNVAGNPEYL 1140  
 Qy 1198 TPOGGAOPHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPSTTFKGT 1241  
 Db 1141 TAQ-----PTCLSSGFNSPALIQKSHQMSLDNPDYQDDFPFKTKNGIFKG-P 1190  
 Qy 1242 TAENPEYLGLDVP 1254  
 Db 1191 TAENAEYLRVAPP 1203  
  
 RESULT 4  
 Q9YH40 PRELIMINARY; PRT; 1165 AA.  
 AC Q9YH40;  
 DT 01-MAY-1999 (T-EMBLrel. 10, Created)  
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
 GN Receptor tyrosine kinase proto-oncogene.  
 GN XMRK.  
 OS Xiphophorus xiphidium.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphia; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 OX NCBI\_TaxID=8086;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RX MEDLINE=98241172; PubMed=9582016;  
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
 RA Altschmidt J., Schartl M.;  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 RT overexpression and mutational alterations";  
 RL Oncogene 16:1681-1690(1998).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Schartl M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; U53471; RAD10500.2; --  
 DR HSSP; P11362; IFGK.  
 DR InterPro; IPR000345; CytC heme bind.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TyrKC; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 2.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.





```
QY 420 PDLVSFQNLQVIRGRILHNGAYSLLTLOGLGISWLGRLSRLSGSLALIHNNTHLCFVHT 479
Db 403 NDLVSFSSUTTTQGRSLKFRFLSMVMWRITPFLSLGRLSRLSREISDGSVYISQNAHLCYRHT 462
QY 480 VPMQDLFRNPH-QALLHTANREDECEVGEGLACHQLCARGHCWGPQTCVNCQSFRLGQ 538
Db 463 VNMVQLFRGSRVRANSLSNRPMACVADGRVCDPLCSDSGCKGPGPDQCLSCRNSRHG 522
QY 539 ECVSECVRLQGLPREYVNRH-CLPCHPCQFQNGSVTCFGEPAQOCVACAHYKPPFCV 597
Db 523 TCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTKASCTGPGACDECMACTKFRDGPYCM 582
QY 598 ARCPQVIRKANSFIGITELPDEEGACOPCINCTHSCVDLDDKGPAPQASPLTSIVA 657
Db 583 SSCPAGVNDGKGL-IFKPNREHCEPCQNCCTQCCSGPGLNDC---LEAARLTSSQ 638
QY 658 VVGILLVVLGVVF-----GILIKRQOKIRKYMRLLELQETELVEPLTPSGAMPNOA 710
Db 639 ITGIALGVPAGLIFCLVFLGLMVLHRLAIRKRAMRYLSSGESFELGP-GEKGTKV 697
QY 711 QMRILKETELRVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRNTPSKANKELDE 770
Db 698 HARILKPSDLRKIKPLGSGVFGTSGFWMPEGETVKIPVAIKTIQDSSGRQTFTEITDH 757
QY 771 AYVMAGVGSPPVRLIGLICTSTVOLVTOLMPYGLLDHVRNRLGSLGODLLNMCQTA 830
Db 758 LLSMGLSDHPYVRLGLICPGCLQVLQSLSHGSLLEHROKTSLOPQRLNMCVQIA 817
QY 831 KMSYLEDLVRLVHRLDAARNLVKSPNHVKITDFGLARLLDDETEYHADGKVKPIKMA 890
Db 818 KGMYYLEEVVHKNLAARNILLKNDYQVOISDYGVADLLYPDKKYYVSETKPIKMA 877
QY 891 LESILRRFTQSDVMSGVTTWELMTFGAKYDGIPIAREIPDLLEKGERLPPOPICTID 950
Db 878 LESILFRYTHQSDVMSGVTTWEMMSFGAEPYASVQPEVPSVLEKGERLSQAICTID 937
QY 951 VYIMVVKCWMIDSECRPRFELVLSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Db 938 VYVMVVKCWMIDENIRPTFKELASDFTRMARDPPYLVIRMEG-----E 981
QY 1011 DDDMGDLVDABEYLVPOQGFPCDPAPGAGMVRHHRSSSTRSGGDLTLGLEPSEEEA 1070
Db 982 DSGMGFL-----GLGDRFATPSLQSPSSWSTSPSQINSVMVMTQLRYD--- 1044
QY 1071 PRSLAPSEGAGSDVFDGLNG-----AAKGLSLFTHDPSPLQ-----RYSDEPT 1117
Db 1009 -----GLGDRFATPSLQSPSSWSTSPSQINSVMVMTQLRYD--- 1044
QY 1118 VPLPSETDGYVAPLTCSPQ- EYVNO-----PDVRPQPSREGPL--PAA 1160
Db 1045 --FAVSQGHIGYLPMSFVDITQLWYQSRSLSSVRLTPDRSAFRSSREAEELCEDGA 1102
QY 1161 RPAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPOGGAAPQPHPPAFSPAFDNL 1220
Db 1103 QCAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1221 YVWQDPPERAGPSTFKGTFAENPE 1247
Db 1123 ---QORKLSTASSPSSFKTWADEDE 1146
```

## RESULT 7

```
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.
AC Q9BIH9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Prterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
```

Anopheles.  
NCBI TaxID=7165;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-SUA;  
RA Lycett G.J.;  
RT "Cloning, expression and localisation of the Anopheles gambiae  
epidermal growth factor receptor."  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ301655; CAC35008.1; .  
DR HSP; P11362; LFGL  
DR InterPro; IPR000345; CysC\_heme\_bind.  
DR InterPro; IPR00494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00669; Pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 7.  
DR SMART; SM00220; S\_TKc; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 4.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW Receptor.  
FT NON\_TER 1 1  
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.8%; Score 2028.5; DB 5; Length 1433;  
Best Local Similarity 32.6%; Pred. No. 3.5e-144;  
Matches 472; Conservative 198; Mismatches 392; Indels 385; Gaps 38;

```
QY 26 CTGTDMLKRLPASPETHLDMLRLHYQGCVVQGNLELTYLPTNASISFLQDIQEVOGYVL 85
Db 1 CIGTNGRMSVPANREHYKRLDRYTNCTYVDGNLEITWQNIITDLNFIQHREVTVGYVL 60
QY 86 IAHQVROVPLORLIVRGTLF-----EDNYALAVLNGDDPLNNTTPTVTGASPGGLREL 140
Db 61 ISLYDLPOVILPRLQIRGRITTFKLNKWEAYGLFV-----SFSHMNTL 104
QY 141 QLRSLTELKGVLIQRNPQLCYQDTILWKDI-FHKNQLALTLDTNRSRACHCSPMC 199
Db 105 ELPLARDILGSGVGFNNYLCHMKSNWBEILLAPQTSMQYTFNFSPVCPCHPSC 164
QY 200 KGSRCWGESSEDCOSLTRTVCAGCA--RCKGPLPTDCHEQCAAGCTGPKHSDCLACLH 257
Db 165 EVG-CWGEAHNCQRFKLNCSPOCSQGRCFGPKPRECHLFCAGCTGTQSDCLACKN 223
QY 258 FNHSGICELHCPALVTYNTDTFESMNPNEGTYTFGASCVTACPNYLSLTDVSGSCTLVCP 317
Db 224 FYDDGVCKQCEPQMIVNPTNYFWEPNPDGKYAYGATCVRKCP-EHLLKDNACVCKCPK 282
QY 318 HNQEVTAEDGTQRCCKSKPCARVCYGLGMQYIKANSKFIGITELEFAGCKKIFGSLAFL 377
Db 283 GKMPQNSE-----CVCKGVCPKTCPGEGI-----VHSDNIG-----NYKDTIIEGSLRL 329
QY 378 PESFDGDPASNT-----APLOPEQLQVFPETLEEITGYLYISAWPDSLPDLSVFNQLQ 429
Db 330 DQSGDFGQVYTNFSGFPRIKIDPDRLEVFSTVKEITGFINIQAHHNFNTLNYFRNLE 389
QY 430 VIRGRILHNGAY-SLTQGLGISWLGRLSRLSGSLALIHNNTHLCFVHTVPMQDLFRN 488
Db 390 VVGRQLKENLFAVYIVKTSLSKLSLELKRNVSGSIVILENSDLCPFVEDIDWSEIKKS 449
QY 489 PHQALLHTANRPEDECVGEGLACHQLCARGHCWGPQTCVNCQSFRLGQSCVCECVLQ 548
Db 450 SDHEVMVQKRNATECHEEGMECSEQCKAGCWKGPEQCLECKRVKYKGLCLDSCK--- 506
```

QY 549 GLPREY-VNARHCLPCHPCQONGSVTCFGEADOCVACAHYKDPFCVACRP----- 601  
Db 507 SLPLRLYSVDSKTCGDCHCEKD-----FCYGNEDNCGSMVYKGRFCVACPTTKHAM 561  
QY 602 --QYKANSKFTGI-----TELPD----- 618  
Db 562 NGTCINCHKTCVGCRCGRPTDIAPDGCISCDKALIGSDAKIERCLMKDESCPDGYSDYVL 621  
QY 619 -EEG----- 621  
Db 622 QEEGPLQLSGRAVCRCKCHPRCKKCTGYGFHEQFCQECTGYKKGECDECPDYANEE 681  
QY 622 --ACQPCPINCT-----HSCVDL-----DD-----KCPAEQ----- 646  
Db 682 TRICLPCHQECRGCHGLGDHHECRNLKLFEGDPYDNATFTVCNCPASHPYKRPQBA 741  
QY 647 -----RASPLTISVSAVVGILVVLGVVFGI---LKRROQKIRKTYM 687  
Db 742 GKIGPYCSADSMQSLRIEPTQVKIVMGSMALILLCVVFGIAFVLSRHKKNKDAVKM 801  
QY 688 RRLQETELVEPLTPSGAMPNQOMRILKETELURKVKVLGSGAGFTYKGIWIPDGENVK 747  
Db 802 TMALAGCEDSEPLRPSNPGNLTKLRIKEAETIRRGVGLMGAFGRVFKGVMPGESVK 861  
QY 748 IPVAIKVLRENTSPKANKETLDEAYVMAGVSPVSRLLGICLTSTVQLVTOLMPYGCLL 807  
Db 862 IPVAIKVLEMSGESSEKFELEAYIMASVEHNLKLLKLAACVMTSQMLITQMLPGCLL 921  
QY 808 DHVRENRGLSGDNLNWCQIAKGMYSLEDRVLRDLAARNVLKPSNVHVKITDFGLA 867  
Db 922 DYVRNKKDKIGSKALLNWSQIARGWAYLEERLVRDLAARNVLQTPSCVKITVFGLA 981  
QY 868 RLIDIDETEYHAGGKVPKIKMALESILRRRFTHQSDVWSYGVTVVWELMTFGAKPYDGP 927  
Db 982 KLDFDSDSEYRAAGGKMPIKWLAECTIRHRVFTSKSDVWAFGITIWELLTYGARPYENVP 1041  
QY 928 AREIPDLKEGERLPPOPICTIDVYIMVKWIMDSERPRFELVSEFSERMARDPQFV 987  
Db 1042 AKDVPELTIEIGHLPQPDICSLDYCILSCWLDADARTFKQLAETFAEKARDPGRYL 1101  
QY 988 VIONEDLGPAASPLDSTFYRSLLEDDMDGLV----- 1018  
Db 1102 MI-----PGDKFMRPLSYTNQDEKDLIRTLAPVMAAAAAAAGASNDVPSTIA 1152  
QY 1019 DAEYLVPOQGFPCPDAPCAGGMVHRHRSSTRSGGDLTLGLPSEEAPRS----- 1073  
Db 1153 ETDEYLPQKTRPSIMULPGPSA-----VEPS-DEMPKSLRYCK 1188  
QY 1074 -PLAP---SEGAGSDVFDGDLGMAAGLQSLPHDPSPLQRYSEDPTVPLPSETDGYVA 1129  
Db 1189 DPLKPDDETGHKEV-----GVGGIR-----LNLPLDEDDYLM 1222  
QY 1130 PLTCSQPOEYVNDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVKDVPFAGG 1189  
Db 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGVPA 1243  
QY 1190 AVENPEYL-----TPQGAAPQPPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGT 1240  
Db 1244 SVDNPEYLMGSTQAIAGLAGSGMG--PHTPP-----PPNTNGM 1280  
QY 1241 PTAENPE 1247  
Db 1281 PTHQHSQ 1287  
RESULT 8  
Q9UK79  
ID Q9UK79 PRELIMINARY; PRT; 419 AA.  
AC Q9UK79;  
DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Herstatic.

GN HER-2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99415951; PubMed=10485918;  
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;  
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted  
RT autoinhibitor.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;  
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF177763; RAD56009.2; -.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR002174; Furin-Like.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L domain; 1.  
DR SMART; SM00261; FU; 1.  
SQ SEQUENCE 419 AA; 45472 MW; FFC1BE347E2D030C CRC64;  
Query Match 27.5%; Score 1871; DB 4; Length 419;  
Best Local Similarity 98.8%; Pred. No. 4.5e-133;  
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASSTVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60  
QY 61 ELYLPTNASLSFLQDIEQVGVLIHQAHOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
Db 61 ELYLPTNASLSFLQDIEQVGVLIHQAHOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180  
Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDTILWKDIFHKNNOLA 180  
QY 181 LTLIDNTRGRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGCARGKPLPTDCCHQEC 240  
Db 181 LTLIDNTRGRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGCARGKPLPTDCCHQEC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRTYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGL 345  
Db 301 YNYLSTDVGSCTLVCPHNOEVTAEQGTORCEKSKPCARVCYGL 345  
RESULT 9  
Q8R2X1  
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.  
AC Q8R2X1;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 40.2 kDa protein.  
OS Mus musculus (Mouse)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC027080; AAH27080.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;



Query Match 25.6%; Score 1739; DB 11; Length 367;  
Best Local Similarity 88.0%; Pred. No. 3.6e-123;  
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPDPPICT 948  
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPDPPICT 60

QY 949 IDVYIMVWKWMDIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008  
DB 61 IDVYIMVWKWMDIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1009 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 1068  
DB 121 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 180

QY 1069 EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDPTVLPSETDGYV 1128  
DB 181 EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDPTVLPSETDGYV 240

QY 1129 APLTCSQPQEVYNQPDVPRQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 1188  
DB 241 APLTCSQPQEVYNQPDVPRQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 300

QY 1189 GAVENPEYLTPOGGAAPQHPFAPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 1248  
DB 301 GAVENPEYLTPOGGAAPQHPFAPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 360

QY 1249 LGLDVVP 1255  
DB 361 LGLDVVP 367

## RESULT 10

Q8WYV0  
ID Q8WYV0 PRELIMINARY; PRT; 412 AA.  
AC Q8WYV0;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Hypothetical 44.7 kDa protein.  
GN PP3659.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
RA Wan D.F., Gu J.R.;  
RT "Novel human cDNA clones with function of inhibiting cancer cell  
RT growth";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF318349; AAL55856.1; -  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match 25.0%; Score 1697.5; DB 4; Length 412;  
Best Local Similarity 80.5%; Pred. No. 5.8e-120;  
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPDPPICT 948  
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPDPPICT 60

DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPDPPICT 60

QY 949 IDVYIMVWKWMDIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1008  
DB 61 IDVYIMVWKWMDIDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1009 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 1068  
DB 121 LEDDDMGDLVDABEYLVPQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLPSEE 180

QY 1069 EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDPTVLPSETDGYV 1128  
DB 181 EAPRSLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDPTVLPSETDGYV 240

QY 1129 APLTCSQPQEVYNQPDVPRQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 1188  
DB 241 APLTCSQPQEVYNQPDVPRQPPSPREGPLPAAPAGATLERAKTLPKNGVVKDVFAFG 300

QY 1189 GAVENPEYLTPOGGAAPQHPFAPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 1248  
DB 301 GAVENPEYLTPOGGAAPQHPFAPAFDNLVWDODPPERCAPSTFKGTPTAENPEY 360

QY 1221 YWYD-QDPPER-----GAPSTFKGTPTAEN 1245  
DB 361 WYVTCQPEGVRRSPDVSSSGSREGLTSAKIKRWEGPPTTSRGTCARN 410

## RESULT 11

Q86712  
ID Q86712 PRELIMINARY; PRT; 729 AA.  
AC Q86712;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Polyprotein.  
GN POLYPROTEIN.  
OS Avian rous-associated virus type 1.  
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
OX NCBI\_TaxID=11950;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=94203659; PubMed=8152791;  
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
RA Johnson A., Beug H.;  
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
RT different transforming capacities";  
RL Oncogene 9:1307-1320(1994).  
DR EMBL; S69372; AAC60725.1; -  
DR HSSP; P03322; 1A6S.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR004028; Retro\_M.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF02813; Retro\_M; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914BE1D63 CRC64;

Query Match 24.8%; Score 1690; DB 15; Length 729;  
Best Local Similarity 54.2%; Pred. No. 5.1e-119;  
Matches 354; Conservative 75; Mismatches 122; Indels 102; Gaps 15;

QY 569 PQNGSVTCFGEADOCVACAHYKDPFCVARGCPQVTKANSKFITELPDEGACQCPPI 628  
DB 141 PEETATPKTGP--DHCMKCNHFDHCHVCVACAGVLGENDTL-VWKYADANAVCOLCHP 197

QY 629 NCHTSCVDLDDKCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKTYM 687  
DB 1 NCHTSCVDLDDKCPAEQASPLTSIVSAVV-GILLVVVLGVVFGILIKRQOKIRKTYM 687

```
Db 198 NCTRGCKGPGLEGCP---NGSKTPSIAAGVVGGLCLVWVGLGIGLYLRRR-HIVRKRTL 253
Qy 688 RLLQETELVEPLTPSCAMPNQAMRILKETELRKVKVLGSGAGTGYKGIWIPDGENVK 747
Db 254 RLLQERELVEPLTPSGEAPNQHLRIKETEFKKVKVLGSGAGTGYKGLWIPGEGBKVK 313
Qy 748 IPVAIKVLRNTPSKANKELIDEA VMAGVSPVSRLLGICLTSTVQLVTQMLPYGCLL 807
Db 314 IPVAIKELREATSPKANKELIDEA VMASVDNPRVCRLLGICLTSTVQLITQMLPYGCLL 373
Qy 808 DHVRENRLGSGDOLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKITDFGLA 867
Db 374 DYIREHNDKICSGVLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKITDFGLA 433
Qy 868 RLIDIDETEHADGKVPKIMWALLESILRRFTHQSDVMSVGVTVWELMTFGAKPYDGIP 927
Db 434 KLGADEKEYHAEGRKVPKIMWALLESILHRIYTHQSDVMSVGVTVWELMTFGSKPYDGIP 493
Qy 928 AREIPDLLEKGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFV 987
Db 494 ASEISSVLEKGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFV 553
Qy 988 VIQ-NEDLGASPLDSTFYRSLLDEDDMDGLVDAAEYLVPOQGFCDPAPGAGMVHHR 1046
Db 554 VIOGDERMHLPSPTDSKPYRTLMEEDMEDIVDAEYLVPHQGF-----598
Qy 1047 HRSSTSTRSGGDLTLGLEPSEEA PRSPL-----APSEGAGSDVFDGLGMGAAGKLSL 1101
Db 599 -NSPST-----SRTPLLSLSATSNNSATNCID-----RNGQGH 631
Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVQPDVVRPPSPREGPLPA 1159
Db 632 PVREDSFVQRYSDPTGNFLSEIDDGFL-----PAPEYVQ--LMPKKPS-----675
Qy 1160 ARPAGATLERAKTUSPQNGVVKDVF-----APFGAVENPEYL 1197
Db 676 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 715

RESULT 12
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynocheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON_TER 1
```

```
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;
Query Match 24.8%; Score 1688; DB 15; Length 567;
Best Local Similarity 54.8%; Pred. No. 4.9e-119;
Matches 353; Conservative 73; Mismatches 116; Indels 102; Gaps 15;

Qy 578 GPBACQVACAHYKDPFCVAPCPQYIKANSKFKIGITELPDEEGACOPCPINCTHSCVDL 637
Db 1 GP--DHCMKAHFIDGPHCVKACPAVGLGENDTL-VWKYADANAVCQLCHPNCCTRGCKP 57
Qy 638 DDGCPAPQASPLTSTIVSAV--GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLCLVWVGLGIGLYLRRR-HIVRKRTLRLQLEREL 113
Qy 697 VEPLTSGAMPNQAORILKETELRKVKVLGSGAGTGYKGIWIPDGENVKIPVAIKVLR 756
Db 114 VEPLTSGEAPNQAORILKETELRKVKVLGSGAGTGYKGLWIPGEGBKVKIPVAIKELR 173
Qy 757 ENTSPKANKELIDEA VMAGVSPVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRR 816
Db 174 EATSPKANKELIDEA VMASVDNPRVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKN 233
Qy 817 LGSODLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKITDFGLARLLDIDETE 876
Db 234 IGSQYLLNWCQIAKMSYLEDVRLVHRDLAARNVLKSPNHNKITDFGLAKLGADEKE 293
Qy 877 YHAGGKVPKIMWALLESILRRFTHQSDVMSVGVTVWELMTFGAKPYDGIPARIPDLLE 936
Db 294 YHAGGKVPKIMWALLESILHRIYTHQSDVMSVGVTVWELMTFGSKPYDGIPASEISSV 353
Qy 937 KGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFVIO--NEDLG 995
Db 354 KGERLPQPPICITIDVYIMVWKMWIDSECRPRFRELSEFSRMARDPQRFVIOGDERMH 413
Qy 996 PASPLDSTFYRSLLDEDDMDGLVDAAEYLVPOQGFCDPAPGAGMVHHRHRSSTSTRSG 1055
Db 414 LPSPTDSKPYRTLMEEDMEDIVDAEYLVPHQGF-----NSPST---454
2y 1056 GGDLTLCLEPSEEA PRSPL-----APSEGAGSDVFDGLGMGAAGKLSLTHDPSPLO 1110
Db 455 -----SRTPLLSLSATSNNSATNCID-----RNGQGHVPVREDSFVQ 491
Qy 1111 RYSEDPTVPLPSET--DGYVAPLTCSPQPEYVQPDVVRPPSPREGPLPAARPAAGATLE 1168
Db 492 RYSSDPTGNFLSEIDDGFL-----PAPEYVQ--LMPKKPS-----526
Qy 1169 RAKTUSPQNGVVKDVF-----APFGAVENPEYL 1197
Db 527 -----TAMVQNOIYNNISLTAISKLPMDSRVQNSHSTAVDNPEYL 566

RESULT 13
Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
GN GAG, V-ERB-A, V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1969616;
RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24(1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
```

```
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stchrmn_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00219; Tyrc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;

Query Match 23.9%; Score 1623.5; DB 15; Length 962;
Best Local Similarity 51.2%; Pred. No. 8.2e-114;
Matches 354; Conservative 73; Mismatches 146; Indels 119; Gaps 18;

QY 541 VECRVQLPLRE-YVNAR-HCLP-----CHPEQC 568
DB 354 IEKQESYLLAFEHVINYKKNIPFWSKLMKVADLRMIGAYHASFRLHMKVCEPTELS 413
QY 569 PONGSVTCFPEADOCVCAHYKDPFCVAPCPQYIKANSKFGITELPDESGACQCPPI 628
DB 414 PGE-----VGP--DHCKMCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHP 465
QY 629 NCTHSCVDLDDGCPAQRASPLTSIVSAV-V-GILLVVVLGVVFGILIKRROOKIRKYTM 687
DB 466 NCTRGCKGPLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRR-HIVKRTL 521
QY 688 RLLQSTELVEPLTSGAMPNQAQWRIKTELKVKVGLGSAFGTVYKGIWIPGENVK 747
DB 522 RLLQERELVEPLTSGEAPNQAHLRIKETEFKKVVLGFGAFGTVYKGLIPEGEKVT 581
QY 748 IPVAIKVLRENTSPKANKEILDEAYVMAGVGSPTYSRLIGICLTSTVOLVTOLMPYGCIL 807
DB 582 IPVAIKELRENTSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQITOLMPYGCIL 641
QY 808 DHVRENRLGSGDILLNWCQIAKGMVSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLA 867
DB 642 DYIREHKDIGNISOYLLNWCQIAKGMVLEERHVRDLAARNVLKTPQHVKITDFGLA 701
QY 868 RLLDIDETEHYADGKVPKIKMALESILRRTHQSDVWSYGVTVWELMTFGAKPYDGP 927
DB 702 KQLGADEKEYHAEGKVPKIKMALESILRIYTHQSDVWSYGVTVWELMTFGSKPYDGP 761
QY 928 AREIPDLLEKGBRLPQPPICTIDVYIMVWKMIIDSECRPRELVSFSESRMARDPQRTV 987
DB 762 ASEISSVLEKGBRLPQPPICTIDVYIMVWKMSGADSRPKRELIAEFSEKWARDPPRYL 821
QY 988 VIQ-NEDLGPASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCPDAPGAGGMVHR 1046
DB 822 VIQDERMHLPSPTDSKFYRLTMEEDMEDI VDAEYLVPHQGF-----866
QY 1047 HRSSTRSGGDLTLGLPSEEAAPRPLAPSEAGSDVDFDGLMGAKGLQSLPTHP 1106
DB 867 -NSPST-----SRTPLLSLSATN-----NSATKCIDRNGGH--898
QY 1107 SPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYVNPQVRPQPPSPREGPLPAAPAGAT 1166
DB 899 -----PVREDGFL-----PAPYVNGQ--LMPKKKSTAMVQNIQYINISLT 936

QY 1167 -LERAKTLSPGKNGVVKVFAFGGAVENPEYL 1197
DB 937 AISKLPMDSRYNQ-----SHSTAVDNPEYL 961

RESULT 14
Q85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Ta34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=88217326; PubMed=28971102;
RX Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278(1987).
DR EMBL; X06943; CAA30024.1; --
DR HSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 60899 MW; 140DC8CCA0F8AF4 CRC64;

Query Match 23.7%; Score 1615; DB 15; Length 545;
Best Local Similarity 54.3%; Pred. No. 1.5e-113;
Matches 341; Conservative 70; Mismatches 125; Indels 92; Gaps 15;

QY 578 GPEADQCVACAHYKDPFCVAPCPQYIKANSKFGITELPDESGACQCPINCTHSCVDL 637
DB 1 GP--DHCKMCAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPCTRGCKGP 57
QY 638 DDKGCPAQRASPLTSIVSAV-V-GILLVVVLGVVFGILIKRROOKIRKYTMRLLOSTEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGIGLYLRR-HIVKRTLRELLQEREL 113
QY 697 VEPLTPSGAMPNQAQWRIKTELKVKVGLGSAFGTVYKGIWIPGENVKIPVAIKVL 756
DB 114 VEPLTPSGEAPNQAHLRIKETEFKKVVLGFGAFGTVYKGLIPEGEKVTIPVAIKEL 173
QY 757 ENTSPKANKEILDEAYVMAGVGSPTYSRLIGICLTSTVOLVTOLMPYGCILLDHVRENRL 816
DB 174 EATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQITOLMPYGCILLDVIHEKDN 233
QY 817 LGSQDLNWCQIAKGMVSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETE 876
DB 234 LGSQDLNWCQIAKGMVLEERHVRDLAARNVLKTPQDVKITDFGLAKQLGADKE 293
QY 877 YHADGKVPKIKMALESILRRTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLE 936
DB 294 YHAEKGKVPKIKMALESILRIYTHQSDVWSYGVTVWELMTFGSKPYDGI PASEISSVLE 353
QY 937 KGERLPQPPICTIDVYIMVWKMIIDSECRPRELVSFSESRMARDPQRFVVIQ-NEDLG 995
DB 354 KGERLPQPPICTIDVYIMVWKMSDADSRPKRELIAEFSEKWARDPPRYLVYIQDERMH 413
QY 996 PASPLDSTFYRSLLEDMDGLVDABEYLVPOQGFPCPDAPGAGGMVHRHRSSTRSG 1055
DB 414 LPSPDTSKFYRLTMEEDMEDI VDAEYLVPHQGF-----NSPST---454
```

QY 1056 GGDLTGLPSEEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQ 1110  
DB 455 -----SRTPLLSLSATSNNATNCIDRNGG-----H----- 481  
QY 1111 RYSEDTFVLPSTGTVAPLTCSPQPEYVNPQDVRPQPSREGPLPAARAGAT-LE 1169  
DB 482 -----PVREDGFL-----PAPEYVNO--LMPKBPSTAMVQIYNYISLTAISK 523  
QY 1170 AKTLPKKGKGVKDVFAFGGAVENPEYL 1197  
DB 524 LPMDSRYON-----SHSTAVDNPEYL 544

## RESULT 15

Q9WVF5 PRELIMINARY; PRT; 655 AA.  
AC Q9WVF5;  
DT 01-NOV-1999 (T-EMBLrel. 12, Created)  
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).  
DE EGFR.  
GN Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=LIVER;  
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,  
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
RA Mailhe N.J.;  
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;  
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
RA Mailhe N.J.;  
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=LIVER;  
RX MEDLINE=21085660; PubMed=11217851;  
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehii Y.,  
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,  
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,  
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,  
RA Schriml L.M., Straubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
RA Hayashizaki Y.;  
RT "Functional annotation of a full-length mouse cDNA collection";  
RL Nature 409:685-690(2001).  
DR EMBL; AF124513; AAD44149.1; -;  
DR EMBL; AF275366; AAG28047.1; -;

DR EMBL; AF275364; AAG28047.1; JOINED.  
DR EMBL; AF275365; AAG28047.1; JOINED.  
DR EMBL; AK004944; BAB23688.1; -;  
DR EMBL; AK004883; BAB23641.1; -;  
DR EMBL; AK004911; BAB23662.1; -;  
DR MGD; MGI:95294; Egr.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR002174; Furin-like.  
DR Pfam; PF00757; Furin-like, 1.  
DR Pfam; PF01030; Recept\_L\_domain; 2.  
DR SMART; SM00261; FU; 3.  
DR Receptor.  
SQ SEQUENCE 655 AA; 72906 MW; 6834063B1BC928CB CRC64;  
Query Match 21.7%; Score 1478.5; DB 11; Length 655;  
Best Local Similarity 44.2%; Pred. No. 4.3e-103;  
Matches 283; Conservative 94; Mismatches 238; Indels 25; Gaps 9;  
QY 11 LLLALLPPGAA--STOVCTGTDMLRLPASPETHLDMLRLHYQCQVVOGNLEITYPTN 68  
DB 14 LLTALCAAGALBEKKVCOQTSNRLTQGTGFEDHFLSLQRMYNNEVVLGNLEITYVQRN 73  
QY 69 ASLSFLQDIOEVQGYVLIHNNQVRQVPLQRLRIVRGTLQFDPEDNYVALAVLQNDPLNNTTP 128  
DB 74 YDLSFLKTIQEVAGYVLIALLNTVERIPLENLQIRGNALYENTYVALAILSN----- 124  
QY 129 VTGASPGRLRELQRLSITELKGGVLIQRPOLCYODTILWKDI----FHKNNOLALTLI 184  
DB 125 -YGTNRTGLRELPMRLQELIIGAVRFSNNPILCNMDTIQWRDVIQVNFMSNMSMDL--- 180  
QY 185 DTVNRSRACHPCSPMKGSRGCGSESSDCQLSTRVTCAGGCA-RCKGPLPTDCCHCEQCAAG 243  
DB 181 -QSHPSCKPKDCPCPNCGSCMGGEENCOKLTIXIAQCCSHRCGRSPSDCCCHNQCAAG 239  
QY 244 CTGPKSHDDCLACLHFNHSGICELHCPALVYNTDTTFESMNPBGRYTFGASCVTACPNY 303  
DB 240 CTGPRESDECLVCQKQFDEATCKDTPPLMLNPTTYQMDVNPBGRYTFGASCVTACPNY 299  
QY 304 LSTDVSGCTLVCLPHNQEVTAEDGTORCEKCPKPCARVCYGLGQYIKANSKFIGITELE 363  
DB 300 VVTDHSGSVRACGPDYEV-EEDGIRKCKKCDGPRKVCNGIGIGEFK-DTSLSNATNIK 357  
QY 364 -FAGCKKIFGLAFSPESFDGDPASNTAPLQEPQLQVFTLEETLYLISAWPDSLPDL 422  
DB 358 HFKYCTAISGDLHLPAFKGDSFTRTPPLDPRELEILKTVKEITGFLLIQAMPDNDL 417  
QY 423 SVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRSLRELSGGLALIHHTHLCFVHTVPW 482  
DB 418 HAFENLEIRGRTRKQHGQFSLAVVGLNITSLRLSLKEISDGDVILSGNRNLCYANTINW 477  
QY 483 DQLPRPHQALLHTANRPEDECGEGLACHOLCARGHCGWPGPTQCVCNCSQFLRGECVE 542  
DB 478 KKLFGTPNQTKTMMNRAEKDCAVNHVNCPLCSSBECGCGPEPRDCVSCQNVSRGECVE 537  
QY 543 ECRVLOGLPREYVNAHRLCHPCECPQNGSVTCFGEADQCACAHYKDPFPFCVARCPQ 602  
DB 538 KCVILEGEPREFVENSEICQHEPECLPQAMNITTCGRPDNCIOCAHYIDGPHCVKTCQA 597  
QY 603 YIKANSKFITITELPEEGACQCPINCTHSCVDLDDKGC 642  
DB 598 GINGENNTL-VKMYADANNVCHLCHANCTYGCAGPGLOGC 636

Search completed: July 22, 2003, 09:01:33  
Job time : 53.3575 secs



CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC -1- RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M11767; AAA35808.1; -  
 CC EMBL; M11761; AAA35808.1; JOINED.  
 CC EMBL; M11762; AAA35808.1; JOINED.  
 CC EMBL; M11763; AAA35808.1; JOINED.  
 CC EMBL; M11764; AAA35808.1; JOINED.  
 CC EMBL; M11765; AAA35808.1; JOINED.  
 CC EMBL; M11766; AAA35808.1; JOINED.  
 CC EMBL; M11730; AAA75493.1; -  
 CC EMBL; M12036; AAA35978.1; -  
 CC EMBL; X03363; CAA27060.1; -  
 CC PIR; A25491; A25491.  
 CC PIR; A24571; A24571.  
 CC HSP; P11362; IFGK.  
 CC GENE; HGNC:3430; ERBB2.  
 CC MIM; 164870; -  
 CC InterPro; IPR000494; EGFR\_L domain.  
 CC InterPro; IPR000719; Euk\_pkinase.  
 CC InterPro; IPR002174; Furin-like.  
 CC InterPro; IPR001245; Tyr\_pkinase.  
 CC InterPro; IPR004019; YLP\_motif.  
 CC Pfam; PF00069; pkinase; 1.  
 CC Pfam; PF00757; Furin-like; 1.  
 CC Pfam; PF01030; Recep\_L\_domain; 2.  
 CC Pfam; PF02757; YLP\_2\_domain; 1.  
 CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC SMART; SM00219; FY; 3.  
 CC SMART; SM00219; TyrKc; 1.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE; PSS0011; PROTEIN\_KINASE\_DOM; 1.  
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Polymorphism.  
 CC SIGNAL 1 21  
 CC CHAIN 22 1255  
 CC DOMAIN 22 652  
 CC TRANSMEM 653 675  
 CC DOMAIN 676 1255  
 CC DOMAIN 720 987  
 CC NP BIND 726 734  
 CC BINDING 753 753  
 CC ACT SITE 845 845  
 CC DISULFID 195 204  
 CC DISULFID 199 212  
 CC DISULFID 220 227  
 CC DISULFID 224 235  
 CC DISULFID 236 244  
 CC DISULFID 240 252  
 CC DISULFID 255 264  
 CC DISULFID 268 295  
 CC DISULFID 299 311  
 CC DISULFID 315 331  
 CC DISULFID 334 338  
 CC DISULFID 511 520  
 CC DISULFID 515 528  
 CC DISULFID 531 540  
 CC DISULFID 544 560

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARIANT 654 654 I -> V.  
 FT VARIANT 655 655 /FTid=VAR\_004077.  
 FT VARIANT 1170 1170 I -> V.  
 FT CONFLICT 1255 AA; 137909 MW; 39E9DFA04DCF962 CRC64;  
 SQ SEQUENCE 1255 AA; 137909 MW; 39E9DFA04DCF962 CRC64;  
 Query Match 97.8%; Score 6651; DB 1; Length 1255;  
 Best Local Similarity 97.6%; Pred. No. 0;  
 Matches 1225; Conservative 10; Mismatches 20; Indels 0; Gaps 0;  
 Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKLRLPASPETHLDMLRHLRYGCCVQVGNL 60  
 Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKLRLPASPETHLDMLRHLRYGCCVQVGNL 60  
 Qy 61 ELTYLPTNASLSFLQDIOEQVGYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120  
 Db 61 ELTYLPTNASLSFLQDIOEQVGYLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120  
 Qy 121 DPLNNTTPVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180  
 Db 121 DPLNNTTPVTGASPGGLREQLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKNQLA 180  
 Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWSESDCSLTTRVCAGGCARCKGRLPTCCHEQC 240  
 Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSESDCSLTTRVCAGGCARCKGRLPTCCHEQC 240  
 Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTNTDTFESMPNPEGRYTFGASCVTACP 300  
 Qy 301 YNYLSTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360  
 Db 301 YNYLSTDVSGCTLVCPPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360  
 Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETGYLYISAWPDSL 420  
 Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETGYLYISAWPDSL 420  
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLGLALHNNTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLGSLGLALHNNTHLCFVHTV 480  
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPQPTQCVNCSQFLRGQEC 540  
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPQPTQCVNCSQFLRGQEC 540  
 Qy 541 VECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600  
 Db 541 VECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFCVARC 600  
 Qy 601 PQVIKANSKFITIGITLPEDEGACQPCPINCTHSCVDLDDKGPAPQASPLTSTVSVAWG 660  
 Db 601 PSGVKPDLSPYMPWKPFDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSTVSVAWG 660  
 Qy 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Db 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720

```

Qy 721 RKVKVLSGARGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVAGVGP 780
Db 721 RKVKVLSGARGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVAGVGP 780
Qy 781 YVSRLLGLICTSTVOLVTQMLPYGCLLDHVRNENRGLSQDQLLNWCQIAKMSYLEDDR 840
Db 781 YVSRLLGLICTSTVOLVTQMLPYGCLLDHVRNENRGLSQDQLLNWCQIAKMSYLEDDR 840
Qy 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900
Db 841 LVHRDLAARNVVKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKWM 960
Qy 961 IDSECRPRFELVSEFMRMDPQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPRFELVSEFMRMDPQRFVVIQNEIDLGASPLDSTFYRSLLEDMDGLVDA 1020
Qy 1021 EEVLVPOOGFCFDPAPGAGGWHHRSSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Db 1021 EEVLVPOOGFCFDPAPGAGGWHHRSSSTRSGGDLTLGLPSEEEAPRSPAPSEG 1080
Qy 1081 AGSDVDFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVDFDGLGMAKGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKOVFAFGGAVENPEYLTPO 1200
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGNVGVKOVFAFGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTPKGTPPTAENPEYGLDVPV 1255
Db 1201 GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTPKGTPPTAENPEYGLDVPV 1255

RESULT 2
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (epidermal growth factor
DE receptor-related protein).
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The new oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofth F.J., Doak D.G., Mulvey D.,

```

```

RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-
RL onogenic and oncogenic forms of the neu protein.";
EMBO J. 11:43-48(1992).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTOUGHOUT NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 757 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.

```

```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 6129264583011402 CRC64;

Query Match 86.2%; Score 5862; DB 1; Length 1257;
Best Local Similarity 86.1%; Pred. No. 9,46-308;
Matches 1082; Conservative 56; Mismatches 117; Indels 2; Gaps 2;

QY 1 MELAAACRWGLLLALLPAGASTOVCTGTDKMLRLPASPETHLDMLRLHLYQGCGVQGNL 60
DB 1 MELAAACRWGLLLALLPAGTAGTQCTGTDKMLRLPASPETHLDMLRLHLYQGCGVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEOGVYVLIAHNOVROVPLQRLRIVRGTLFEDNYALAVLDNG 120
DB 61 ELTYVPAVASLSFLQDIQEOGVYVLIAHNOVQVPLQRLRIVRGTLFEDNYALAVLDNR 120
QY 121 DPLNNTTPTVT-GASPGGLRELQLRLSTEILKGGVLIQIORNPOLCYQDTILMKDIFPHKNQL 179
DB 121 DPQDNVAASTPGRTPGELRELQLRLSTEILKGGVLIQIORNPOLCYQDMVLWKDVKRKNQL 180
QY 180 ALTLIDNRSRACHPCSPMKSGSCWGESSEDCSLRTVCAGCARCKGRLPTDCCHEQ 239
DB 181 APVIDIDNRSRACHPCSPMKSGSCWGESSEDCSLRTVCAGCARCKGRLPTDCCHEQ 240
QY 240 CAAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAC 299
DB 241 CAAGCTGPKGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTTC 300
QY 300 PYNVLTSDVSGCTLVCPHNOEVTAEADGTQRCCKSPCARVCYGLGQYIKANSKFIGI 359
DB 301 PYNVLTSEVSGCTLVCPHNOEVTAEADGTQRCCKSPCARVCYGLGMEHLRGARAITSD 360
QY 360 TELEFACKTIFGSLAFELPESFDGDPASNTAPLOEQLQVPELLEETGYLYISAWDSL 419
DB 361 NVOEFDGCKKIFGSLAFELPESFDGDPSSGIAPLPEQLQVPELLEETGYLYISAWPDSL 420
QY 420 PDLVSFQNLQVIRGILHNGAYSLSLQGLGTSWGLSLRLSRELGSGLALIHNTLHLCFVHT 479
DB 421 RDLVSFQNLRIIRGILHNGAYSLSLQGLGHSGLSLRLSRELGSGLALIHNAHLCFVHT 480
QY 480 VPMQDLFRNPHQALLHTANPEDE-CVGEGLACHQLCARGHCWGPPTQCVCNCSQFLRGQ 538
DB 481 VPMQDLFRNPHQALLHSGNRPEDLGVSSGLVCNLSLCAHGCWGPPTQCVCNCSHFLRGQ 540
QY 539 ECVEECRVQLQPLREYNARHCLPCHPEQCPONGSVTCFGEADQCVACAHYKDPPECVA 598
DB 541 ECVEECRVWGLPREYVSDKRLPCHPEQCPQNSSETCFGEADQCAACHYKDDSSSCVA 600
QY 599 RCPQYIKANSKFIGITELPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIVSAV 658
DB 601 RCPSGVKPDLSSYMPIKYPDEEGICQPCINCTHSCVDLDERGCPAQASPLTSIVSAV 660
QY 659 VGILLVVVLGVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKET 718
DB 661 VGILLFVLVVVGVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKET 720
QY 719 ELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAVVMAGVG 778
DB 721 ELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAVVMAGVG 780
QY 779 SPVSVRLGLICLTSTVOLVTQLMPYGCGLLDHVRNRRGLSGQDILLNMCQIAKMSYLE 838
```

```
DB 781 SPVSVRLGLICLTSTVOLVTQLMPYGCGLLDHVRHRRGLSGQDILLNMCQIAKMSYLE 840
QY 839 VRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETETEHADGCKVPIKMALESILRRR 898
DB 841 VRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETETEHADGCKVPIKMALESILRRR 900
QY 899 FTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPOPPICITIDVYMIWVC 958
DB 901 FTHQSDVMSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPOPPICITIDVYMIWVC 960
QY 959 WMIDSECRPRFRELVSFMRMDPQRFVVIQNEQLGPASPLDSTFYRSLLEDDMDGLV 1018
DB 961 WMIDSECRPRFRELVSFMRMDPQRFVVIQNEQLGPASPLDSTFYRSLLEDDMDGLV 1020
QY 1019 DABEYLVPQGGFFCPDPAPGAGGMVHRHRSSTSRGSGDLTLGLSEPEEAPRSLAPS 1078
DB 1021 DABEYLVPQGGFFCPDPPTGCTAHRHRSSTSRGSGDLTLGLSEPEEAPRSLAPS 1080
QY 1079 EGAGSDVDFDGLGMAKGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1138
DB 1081 EGAGSDVDFDGLGMAKGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQE 1140
QY 1139 YVNPQDVRPQPPSPREGPLPAAPAGATLERAKTSLSPKNGVKDVFAGGAVENPEYLT 1198
DB 1141 YVNPQDVRPQPPSPREGPLPAAPAGATLERAKTSLSPKNGVKDVFAGGAVENPEYLV 1200
QY 1199 POGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPPFTFKGTPTAENPEYLGDLV 1255
DB 1201 PREGTAAPHPHPPAFSPAFDNLVYWDQDPPERGAAPPFTFKGTPTAENPEYLGDLV 1257

RESULT 3
ERB2 MESAU
ID ERB2 MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DB (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
EMBL; D16295; BAA03801.1; --  
DR HSP; P11362; IFK.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP motif.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TyKc; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Proto-oncogene; Disease mutation.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 653 675 POTENTIAL.  
FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 158 368 CYS-RICH.  
FT DOMAIN 472 644 CYS-RICH.  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.  
FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 567 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.9%; Score 5846.5; DB 1; Length 1254;  
Best Local Similarity 85.7%; Pred. No. 6.4e-307;  
Matches 1075; Conservative 65; Mismatches 114; Indels 1; Gaps 1;

QY 1 MELAAALCRGLGLALLPPGAASGVCTGTDMLKRLPASPEHLDMLRHLYQGCQVQGNL 60  
DB 1 MELAAACGGLGLALLSPASGVCTGTDMLKRLPASPEHLDLIVRHLYQGCQVQGNL 60

QY 61 ELTYLPTNASLFLQDIOEQVGYVLIHAHNOVRQVPLQRLRIVRGTQLPEDNYALAVLQNG 120  
DB 61 ELTYLPTNASLFLQDIOEQVGYVLIHAHNOVRQVPLQRLRIVRGTQLPEDNYALAVLQNG 120  
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQIRNPQLCYQDITLWKDI FHKQNOLA 180  
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQIRNPQLCYQDITLWKDI FHKQNOLA 180  
QY 181 LTLDITNRSRACHPCSPMKGSCGSGESSDDCSLTTCVAGGCARCKGPLPTCCHEQC 240  
DB 181 LTLDITNRSRACHPCSPMKGSCGSGESSDDCSLTTCVAGGCARCKGPLPTCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNNOEVTAEADGTQCEKSKSCARVCYGLGMEHLRGARAITSAN 360  
DB 301 YNYLSTDVGSCTLVCPLNNOEVTAEADGTQCEKSKSCARVCYGLGMEHLRGARAITSAN 360  
QY 361 ELFPAGCKTFGSLAFELPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAPWSLP 420  
DB 361 ELFPAGCKTFGSLAFELPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAPWSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISLWGLRLSLRELGLGLALIHNTLHCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLGISLWGLRLSLRELGLGLALIHNTLHCFVHTV 480  
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSQFLRGQEC 540  
DB 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSQFLRGQEC 540  
QY 541 VEESCRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVACR 600  
DB 541 VEESCRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVACR 600  
QY 601 POYIKANSKFIGITELPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVG 660  
DB 601 POYIKANSKFIGITELPDEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSTVSAVG 660  
QY 661 ILLVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQMRILKETEL 720  
DB 661 ILLVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQMRILKETEL 720  
QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKI PVAIKVLRNTPSKANKEILDEAYVWAGVSP 780  
DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKI PVAIKVLRNTPSKANKEILDEAYVWAGVSP 780  
QY 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGSODLLNWCQIAKMSYLSQV 840  
DB 781 YVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGSODLLNWCQIAKMSYLSQV 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRRT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKWMALESILRRRT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMWKCM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMWKCM 960  
QY 961 IDSECRPRFELVSEFMRARDPQRFVWQNEQDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
DB 961 IDSECRPRFELVSEFMRARDPQRFVWQNEQDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
QY 1021 EYLVLPQOQFFCPDPAPGAGGMVHRRSSSTSGGGDLTLGLEPSEEEAAPSPLAPSEG 1080  
DB 1021 EYLVLPQOQFFCPDPAPGAGGMVHRRSSSTSGGGDLTLGLEPSEEEAAPSPLAPSEG 1080  
QY 1081 AGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSDTGTYVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPVPLPSDTGTYVAPLTCSPQPEYV 1140

```
Oy 1141 NQPDVRPQPPSPREGPLPAARPAATLERAKTLSPGKGVVKKDVFAGGAVENPEYLTPQ 1200
Db 1141 NQPEVRQPPLTPEGPLPPVRPAGATLERPKTLSPGKGVVKKDVFTEGGAVENPEYLVR 1200
Oy 1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPERGAPSTFKGTPTAENPEYLGLDVVP 1255
Db 1201 GGSASQPH-PPALCPAFDNLYYWDQPPSERGSPNTPEGTPTAENPEYLGLDVVP 1254

RESULT 4
EGFR_HUMAN
ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
AC P00533; P06268; Q14225; Q9UMD8; Q9UMG5; Q92795; O00732;
AC Q00688; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JUN-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RA "Human epidermal growth factor receptor cDNA sequence and aberrant
RA expression of the amplified gene in A431 epidermoid carcinoma cells.";
RA Nature 309:418-425 (1984).
RN [2]
SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RA "Possible role of variant RNA transcripts in the regulation of
RA epidermal growth factor receptor expression in human placenta.";
RA Mol. Reprod. Dev. 41:149-156 (1995).
RN [3]
SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RA "A 1.8 kb alternative transcript from the human epidermal growth
RA factor receptor gene encodes a truncated form of the receptor.";
RA Nucleic Acids Res. 24:4050-4056 (1996).
RN [4]
SEQUENCE FROM N.A. (ISOFORM 2).
RX TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RA "Expression of a truncated epidermal growth factor receptor-like
RA protein (TEGFR) in ovarian cancer.";
RA Gynecol. Oncol. 65:36-41 (1997).
RN [5]
SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RX TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RA "Comparative genomic sequence analysis and isolation of human and
RA mouse alternative EGFR transcripts encoding truncated receptor
RA isoforms.";
RA Genomics 71:1-20 (2001).
RN [6]
SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
```

```
RA Maible N.J.;
RA "Human and mouse alternative EGFR transcripts encoding only the
RA extracellular domain of the receptor.";
RA Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RA "Expression cloning of human EGF receptor complementary DNA: gene
RA amplification and three related messenger RNA products in A431
RA cells.";
RA Science 224:843-848 (1984).
RN [8]
SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RA "Human epidermal growth factor receptor cDNA is homologous to a
RA variety of RNAs overproduced in A431 carcinoma cells.";
RA Nature 309:806-810 (1984).
RN [9]
SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RA "Isolation of an evolutionarily conserved epidermal growth factor
RA receptor cDNA from human A431 carcinoma cells.";
RA Biochem. Biophys. Res. Commun. 124:125-132 (1984).
RN [10]
SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RA "The human EGF receptor gene: structure of the 110 kb locus and
RA identification of sequences regulating its transcription.";
RA Oncogene Res. 1:375-396 (1987).
RN [11]
SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RA "Contributory effects of de novo transcription and premature
RA transcript termination in the regulation of human epidermal growth
RA factor receptor proto-oncogene RNA synthesis.";
RA J. Biol. Chem. 266:1746-1753 (1991).
RN [12]
SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RA "Characterization and sequence of the promoter region of the human
RA epidermal growth factor receptor gene.";
RA Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924 (1985).
RN [13]
SEQUENCE OF 540.
RA Kohda D.;
RA Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Morig G., Cohen S.;
RA "ATP-stimulated interaction between epidermal growth factor receptor
RA and supercoiled DNA.";
RA Nature 309:270-273 (1984).
RN [15]
PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RA "All autophosphorylation sites of epidermal growth factor (EGF)
RA receptor and HER2/neu are located in their carboxyl-terminal tails.
RA Identification of a novel site in EGF receptor.";
RA J. Biol. Chem. 264:10667-10671 (1989).
RN [16]
```



```

QY 1141 NOPDVRPQPSREGPLPAARPACATLERAKTLSPGKNGVVKVFAFGGAVENPEYL-TP 1199
Db 1094 NQ-SVPRKPGSVQNPVHYHNOPLNP-----APSRDPHYQD--PHSTAVGNPEYLNLT 1142
QY 1200 QGGAAPOPHPPAFSPAFDNLVYWDQ-----DP-----PERGAPSPTEFGTPTA 1243
Db 1143 Q-----PTCVNSTFDSPAHWAQKSHQISLDNPDYQODFFPKAKPNIGFKGS-TA 1192
QY 1244 ENPEYL 1249
Db 1193 ENAEYL 1198

RESULT 5
EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RN Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RN Genes Dev. 8:399-413(1994).
[5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RN Oncogene 6:673-676(1991).
[6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisinger D.P., Serrero G.;
RN Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS ALPHA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X78987; CAA55587.1; -
CC EMBL; U03425; AAA17899.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSP; P11362; IFGK.
CC MGD; MGI:95294; Egfr.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00361; FU; 3.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL
CC CHAIN 25 1210
CC DOMAIN 25 647
CC TRANSMEM 648 670
CC DOMAIN 671 1210
CC REPEAT 75 300
CC REPEAT 390 600
CC DOMAIN 1028 1071
CC NP BIND 720 728
CC BINDING 747 747
CC ACT SITE 839 839
CC DISULFID 190 199
CC DISULFID 194 207
CC DISULFID 215 223
CC DISULFID 219 231
CC DISULFID 232 240
CC DISULFID 236 248
CC DISULFID 251 260
CC DISULFID 264 291
CC DISULFID 295 307
CC DISULFID 311 326
CC DISULFID 329 333
CC DISULFID 506 515
CC DISULFID 510 523
CC DISULFID 526 535
CC DISULFID 539 555
CC DISULFID 558 571
CC DISULFID 562 579
CC DISULFID 582 591
CC DISULFID 595 617
CC DISULFID 620 628
CC DISULFID 624 636
CC MOD RES 680 680
CC MOD_RES 1092 1092

```



CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
 CC 2. NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NPAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC -!- NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphatase.  
 CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLAVABLE AND  
 CC NONCLAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC DR EMBL; L07868; AAB59446.1; -;  
 CC DR HSPG; P11362; IFGK  
 CC DR Genew; HGNC:3432; ERBB4.  
 CC DR MIM; 600543; -;  
 CC DR InterPro; IPR000494; EGFR\_L\_domain.  
 CC DR InterPro; IPR000719; Euk\_pkinase.  
 CC DR InterPro; IPR002174; Furin-like.  
 CC DR InterPro; IPR001245; Tyr\_pkinase.  
 CC DR InterPro; IPR004019; YLP\_motif.  
 CC DR Pfam; PF00069; pkinase; 1.  
 CC DR Pfam; PF00757; Furin-like; 1.  
 CC DR Pfam; PF01030; Recep\_L\_domain; 2.  
 CC DR Pfam; PF02757; YLP; 2.  
 CC DR ProDom; PD000001; Euk\_pkinase; 1.  
 CC DR SMART; SM00261; FU; 4.  
 CC DR SMART; SM00219; TyrKc; 1.  
 CC DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 CC DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 CC DR PROSITE; PS50011; PROTEIN KINASE\_DOM; 1.  
 CC DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Alternative splicing.  
 CC FT SIGNAL 1 25  
 CC FT CHAIN 26 1308  
 CC FT DOMAIN 26 651  
 CC FT TRANSMEM 652 675  
 CC FT DOMAIN 676 1308  
 CC FT DOMAIN 186 334  
 CC FT DOMAIN 496 633  
 CC FT DOMAIN 718 985  
 CC FT NP BIND 724 732  
 CC FT BINDING 751 751  
 CC FT ACT SITE 843 843  
 CC FT DISULFID 189 197  
 CC FT DISULFID 193 205  
 CC FT DISULFID 213 221  
 CC FT DISULFID 217 229  
 CC FT DISULFID 230 238  
 CC FT DISULFID 234 246  
 CC FT DISULFID 249 258  
 CC FT DISULFID 262 289  
 CC FT DISULFID 293 304

FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED GLCNAC. (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED GLCNAC. (POTENTIAL).  
 FT VARSPLIC 626 648 NGPTSDCIYPPYMGHSTLTPQHA -> IGSSIEDICIGLMD  
 FT FT (IN ISOFORM JM-B).  
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D88761 CRC64;  
 Query Match 43.5%; Score 2956.5; DB 1; Length 1308;  
 Best Local Similarity 45.2%; Pred. No. 1.4e-151;  
 Matches 609; Conservative 181; Mismatches 383; Indels 173; Gaps 28;  
 Qy 9 WGLLLALLPPGAA-----STVCTCTDMKRLRASPETHLDMLRLHLYOCQVVOGNLEITY 64  
 Db 8 WVVVSLVAAGTVQPSDSQSVCACTENKLSLSLEQOYRALRYKYENCEVVMGNLEITS 67  
 Qy 65 LPTNASLFLQDIOEVQGYVLIHNNQVRQVPLQRLRVRGQLFEDNYVALVONGDPLN 124  
 Db 68 IEHRDLSPFLRSVREVTGYVLVALNQRYLPLENLRIRGTLKLYEDRYALAIFLNRYKDG 127  
 Qy 125 NTPVTVGASPGGLRELQRLRSITELKGGVLIQRNPOLCYODTILKWKDIFHKNNQLATLI 184  
 Db 128 NF-----GLOELGLKNLTELNGGVYVDQNKFLCYADTIHWDIIVRNWPSNLTFLV 178  
 Qy 185 DTRNSRACHPCSPMKGSRGWESSEDCQSLTRTVACAGC-ARCKGPLTDCHEQCAAG 243  
 Db 179 STNGSSGGRCHKSCCTG-RCMGPTEHNCQTLTRTVCAEQCDGRCYGPVSDCHRECAGG 237  
 Qy 244 CTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPVNY 303  
 Db 238 CSGPKDTCDFACMNFNDGACVTCQCPOTFVYNPTTFQLEHNFNAKYTYGAFCKVKCPHF 297  
 Qy 304 LSTDVSGCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFICITILE 363  
 Db 298 V-VDSGCVACRCPSSKMEV-EENGIRKMKCPCTDICPKACDGIPTGSLMSAQTVDSSNDK 355  
 Qy 364 FAGCKTIFGSLAFIPESFDGDPASNTAPLOEQVFEETLEEITGYLYISAWPSLPDLS 423  
 Db 356 FINCTKINGNLIFLVGTGIHGDYPYNAIBAIPEKLNVERTVREITGFLNQSWPNMTDFS 415  
 Qy 424 VFQNLQVIRGRIHLHNGAYSILTQGLGISWIGLSRLSRLGSLGLAIHNNTHLCFVHTVPWD 483  
 Db 416 VFSNLVTIGRVLVYSGLSLLILKQOGITSLSQFQSLKEISAGNIYITDNLNLCYVHTINWT 475  
 Qy 484 QLFRNPQALLHTANRDEDCVGEGLACHOLCARGHCWGPCTOCVNCQSLRQECVEE 543  
 Db 476 TLFSTINQRIIVIRDNRAENCTAGBMVCNHLCSDDGCMGPGDQCLSCRRFSRGRICIES 535  
 Qy 544 CRVLQGLPREYVVARHCLPCHPEQCP-QNGSVTCFGEADQCAAHYKDPFPFCVACPCQ 602

```
Db 536 CNLYDGEFRFENGSIQCECDPOCEKEDGLTCHGPGDNCCKSHFKDGNPCVCEKCPD 595
Qy 603 YIKANSKFIIGITELPDEGACQCPINCTHSCVDLDDKGC-----PAPQASPL 651
Db 596 GLOANSFIKYADPDRE--CHPCHPNCQCCNGPTSHDCIYYPWTHGSHLTPQHAR-TPL 652
Qy 652 TSIVSAVV-GILLVVLGVVFGILIKEROQKIRKYMRLLOETELVEPLTPSGAMPNOA 710
Db 653 --TAAGVIGLGLVILVGLTFVAVVRRKSIK-KKRALRRL-ETELVEPLTPSGTAPNOA 708
Qy 711 QMRILKETELRKVKVLGSGAGFYVYGIWIPDGENVKIPVAIKVLRENTSPKANKETLDE 770
Db 709 QLRILKETELRKVKVLGSGAGFYVYGIWIPDGENVKIPVAIKVLRENTSPKANKETLDE 768
Qy 771 AYVAGVGSYVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRLGSGDLLNMCQIA 830
Db 769 ALIMASMDHPLVRLGLVCLSPITQLVTLMPHGLLEYVHEHKNIGSQLLNMCVQIA 828
Qy 831 KGSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMA 890
Db 829 KGMYLEERLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIMA 888
Qy 891 LESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTID 950
Db 889 LECIHYRKPTHQSDVMSYGVYVWELMTFGCKPYDGPAREIPDLLEKGERLPPOPICTID 948
Qy 951 VYMIVKCMWIDSECPREFELVSEFSRMARDPQRFVIONED-LGPASPLDSTFYRSL 1009
Db 949 VYMIVKCMWIDSECPREFELVSEFSRMARDPQRFVIONED-LGPASPLDSTFYRSL 1008
Qy 1010 EDDMDGLDVAEYLVPOGFFCPDPAAGAGVMVHRHRSSTRSGGDLTLGLESEEE 1069
Db 1009 DEEDLEDMDVAEYLVPOGFFCPDPAAGAGVMVHRHRSSTRSGGDLTLGLESEEE 1056
Qy 1070 APRS-----PLAP-SEGAGSDVDFDGLGMAAGKLOS 1100
Db 1057 TWMGNGQFVYRGGFAAQGVSVYRAPTSITIEAPVAQATAEIPDDSCCNGLRKPA 1116
Qy 1101 LPHDPSPLORVSEDITVPLS-----ETDGYVAPLTCSPQPEYVNOQDVVRPQPSPR 1153
Db 1117 PHVQEDSSQRYSDPTVFAPERSPRGELDEGYMTPMRDKPKQEYLNVE----- 1167
Qy 1154 EGPLPAARPAAGATLERAKTLSPCKNGVVKDVFAGGAVENPEVLTQGGNAPOPHPPA- 1212
Db 1168 ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG-----PPKAE 1199
Qy 1213 -----FSPAFNLVYWDQDPPERGA--PPSTF 1237
Db 1200 DEVVNEPLYLNTFANTLGRAEYLNKILSMPEKAKAFDNDPDYWNHSLPRSTLQHPDYL 1259
Qy 1238 KGTPT-----AENPEYL 1249
Db 1260 QEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 7
ERR4_RAT
ID ERR4_RAT STANDARD; PRT; 1308 AA.
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
CN ERR4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC EMBL; AF041838; AAC08899.1; -.
CC EMBL; U52531; AAC53051.1; -.
CC HSSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; YLP_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC Pfam; PF02757; YLP; 2.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 4.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).
```



FT	DOMAIN	186	334	CYS-RICH.	QY	357	IGITELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEEITGVLYISAMP	416
FT	DOMAIN	496	633	CYS-RICH.	Db	349	SSNIDKFNCTKINGNLFLVTVGIHGDPPYNAIDAIDPEKLVNFRVREITGFLNIOTWP	408
FT	DOMAIN	718	985	PROTEIN KINASE.	QY	417	DSLPLDSVFONQVIRGRILHNGAYSLTIGLIGISWGLSLRSLRGLSLALIHNNHLCF	476
FT	NP BIND	724	732	ATP (BY SIMILARITY).	Db	409	PNMTDFSVFSLVTIGRVLVSLLLKQOQITSLQFQSLKEISAGNIYITDINSNLCY	468
FT	ACT SITE	751	751	BY SIMILARITY.	QY	477	VHTVPDQLPFRNPHQALLHTANPEDECEVGEGLACHOLCARGHGWGPGTQVCNCSOFLR	536
FT	DISULFID	189	197	BY SIMILARITY.	Db	469	YHTINMTTFLSTVYQRIIVDRNRAENCTAEGMVCNHLCSNDGCGWGPQOCLSCRRFSR	528
FT	DISULFID	193	205	BY SIMILARITY.	QY	537	GOECVBECEVLOGLPREYVNNARHCLPCHPEQCP-QNGSVTCFGEADQCACAHYKDPFF	595
FT	DISULFID	213	221	BY SIMILARITY.	Db	529	GKICIESCNLYDGEFREFENGSIQVCDSCQEKMEDGLLTCHGPGPNDCTKCSHFQDGN	588
FT	DISULFID	217	229	BY SIMILARITY.	QY	596	CVARCPOYIKANSKFIGITELPDEBEGACQPCPNCNTHSCVDLDDKGC-----PA	644
FT	DISULFID	230	236	BY SIMILARITY.	Db	589	CVEKCPDLQGANSP--IFKYADQDRCHPCNCTQCGNGPTSHDCIYYPWTGHSTLPQ	646
FT	DISULFID	234	246	BY SIMILARITY.	QY	645	EQRASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPS	703
FT	DISULFID	249	258	BY SIMILARITY.	Db	647	HAR-TPL--IAGVIGGLFLVIMALTFAVYVRKSIK-KKRALRREL-ETELVEPLTPS	701
FT	DISULFID	262	289	BY SIMILARITY.	QY	704	GAMPNOAQRILKETELRKVKVLSGAGFVYVYKGIWIPDGENVKIPVAIKVLRNENTSPKA	763
FT	DISULFID	293	304	BY SIMILARITY.	Db	702	GTAPNOAQLRIKETELRKVKVLSGAGFVYVYKGIWIPDGETVKIPVAIKILNETTGPKA	761
FT	DISULFID	308	323	BY SIMILARITY.	QY	764	NKEILDEAYVMAGVSGPYVSRLLIGICLTSTVQLTQVLMFYGCLLDHVRNRRGLSODLL	823
FT	DISULFID	326	330	BY SIMILARITY.	Db	762	NVEFMDEALIMASVDHPHLVRLLVGLVCLSPITQLVTQVLMFYGCLLEYYVHEHKDNIGSOLL	821
FT	DISULFID	503	512	BY SIMILARITY.	QY	824	NWCQITAKGNSYLEDVRLVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGK	883
FT	DISULFID	507	520	BY SIMILARITY.	Db	822	NWCQITAKGNSYLEDVRLVHRDLAARNVLKSPNNHVKITDFGLARLLDIDETEHADGGK	881
FT	DISULFID	523	532	BY SIMILARITY.	QY	884	VPIKMALESILRRRFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPARIPOLLEKGERLPQ	943
FT	DISULFID	536	552	BY SIMILARITY.	Db	882	MPIKMALECIHYRKFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPARIPOLLEKGERLPQ	941
FT	DISULFID	555	569	BY SIMILARITY.	QY	944	PPICTIDVYIMVKCMWIDSECRPRPRELSEFSRMDARDQRFVVIQNEQ-LGPASPLDS	1002
FT	DISULFID	559	577	BY SIMILARITY.	Db	942	PPICTIDVYIMVKCMWIDSECRPRPRELSEFSRMDARDQRFVVIQNEQ-LGPASPLDS	1001
FT	DISULFID	580	589	BY SIMILARITY.	QY	1003	TFYRSLLDDDDGDLVDAEYLVPOQGFCCDP-----APGA	1039
FT	DISULFID	593	614	BY SIMILARITY.	Db	1002	KFFQNLLEDDEEDMDDAEYLVQ-QAFNIPPIYTSRTRIDSNRSEIGHSPPPAYTPMS	1060
FT	DISULFID	617	625	BY SIMILARITY.	QY	1040	GMVHHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPSEAGSDVDFDGLMGCAAGLQ	1099
FT	DISULFID	621	633	BY SIMILARITY.	Db	1061	GSQFVYQDGGFATQCG--NMPYATATSTIPEAPVA--QGATAEMFDDSCCNGTLRKPV	1115
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-)	QY	1100	SLPHTDPSLQRYSESDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRPQPPSP	1152
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-)	Db	1116	VPHVEDSSQRYSDPTVPEAPERNPRAELDEEGYMTPMHDKPKQEYLNPNVE-----	1167
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-)	QY	1153	REGPLPAARAGATLEBAKTLSPCKGKGVKDVAFGCGAVENPEYLTQCGGAPOPHPPA	1212
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-)	Db	1168	ENPFVSRR-----KNGDLQ-----ALDNPEYHSASSG-----PPKA	1198
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. .)	QY	1213	-----FSPAFLNYWDQDPPPERGA--PPST	1236
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. .)	Db	1199	EDEVNEPLYNTFTNALGNAEYKNSLLSVPEKAKAFDNDPDYWNHSLPRSTRLOHPDY	1258
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. .)	QY	1237	FKGTPT-----AENPEYL	1249
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. .)	Db	1259	LQEYSTKYFYKONGRIRPIVAENPEYL	1285
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. .)	QY			
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. .)	Db			
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. .)	QY			
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. .)	Db			
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .)	QY			
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. .)	Db			
FT	CARBOHYD	1062	1062	S -> N (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	QY			
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).	Db			





```

Db      590  IPRCPHGLDGTLL-IWKYADKMGQCPCHQCTQCCSPGLSGRGD-IVSHSLAVG 647
Qy      657  AVVGIILLVVLGVVFGILIKRROOKIRKYTMRLLOETELVEPLTPSGAMPNOAQRILK 716
Db      648  LVSGLLITVIVALLVLLRRRIK-RKRTIRCLLQEKELVEPLTPSGQAPNQAFLRILK 706
Qy      717  ETELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKILLDEAYVMAG 776
Db      707  ETEFKDVRVLSGAFGVYKGLWNPGENIRIPVAIKVLRENTSPKVNQVLEAYVMAS 766
Qy      777  VGSPPVYSRLIGICLTSTVQLVLTQMPYGCILLDHVRENKRLGSGDILNMCQIAKMSYL 836
Db      767  VDRPHVCRLLGICLTSAVQLVLTQMPYGCILLDVVRQHERICGQWLLNMCVQIAKGNYL 826
Qy      837  EDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILR 896
Db      827  EERHLVHRDLAARNVLLKNPHVKITDFGLSKLLTADKEYQADGGKVPDKWMALESILQ 886
Qy      897  RRTTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYMIW 956
Db      887  WTYTHOSDWSYGVTVWELMTFGSKPYDGIPIAREIPDLLEKGERLPPOPICTIEVYMIIL 946
Qy      957  KCMWIDSECRPRELVSERFMRWARDPQRFVWIQNEDLGPASPLDSTFYRSLLEDDEMDGD 1016
Db      947  KCMWIDSECRPRELVSERFMRWARDPQRFVWIQNEDLGPASPLDSTFYRSLLEDDEMDGD 1001
Qy      1017  LVDAEYLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLERPSEERAPRPLA 1076
Db      1002  VVDAEYLVPOQGFCDPAPGAGVHHRSSSTRSGGDLTLGLERPSEERAPRPLA 1024
Qy      1077  PSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDTV-PLPSETDGYAPLTCSP 1135
Db      1025  PPTGH-----PVRENSITLNRISDPTQNALEKDLQGH----- 1056
Qy      1136  QPEYVNPQDVRPOP-----PSPRE-----GPLP-AARPAGATLERAKTLSPKNGVVK 1182
Db      1057  --EYVNPQGETSSRLSDIYNPNVEDLTCGMGPVLSLSQEAETNFSRPEVLNTNQSL-- 1112
Qy      1183  DVFAFGAVENPEYLTQGAARQPHPPPAFSDNLYWDQDPPPERGAPPSTFKGTPT 1242
Db      1113  -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTGNGMFLPA 1149
Qy      1243  AENPEYLG 1250
Db      1150  AENLEYLG 1157

ID_ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;

```

```

RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29366; AAA5790.1; -.
CC EMBL; M34309; AAA5979.1; -.
CC EMBL; S61953; AAB26935.1; -.
CC PIR; A36223; A36223.
CC HSP; P11362; LFQK.
CC GENE; HGNC:3431; ERBB3.
CC MIM; 190151; -.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Recept_L_domain; 1.
CC Pfam; PF01030; Recept_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU_3.
CC SMART; SM00219; TyrcK; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
CC PROSITE; PS50011; PROTEIN KINASE DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
CC SIGNAL 1
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202

```

FT	DISULFID	210	BY SIMILARITY.	218	QY	483	DOLFRNPQALLHTA-NRPEDECVGEGGLACHOLCARGHCWGPGPTQCVNCSOFLRQECV	541
FT	DISULFID	214	BY SIMILARITY.	226	Db	471	TKVLGRPTTEERLDIKHNRPRDRCDVAEGKVCDCPLCSCGGCGGPGGCLSCRNYSRGVCV	530
FT	DISULFID	227	BY SIMILARITY.	235	QY	542	EECRVLOGLPREYVYVNRHCLPCHPCQONGSVTCFPGPEADOCVACAHYKDPFPCVRCPC	601
FT	DISULFID	231	BY SIMILARITY.	243	Db	531	THCNFLNGEPREFEAHEAEFCFCHPCQPMEGTATCNGSGSDTCAOCHAFRDPCHCVSSCP	590
FT	DISULFID	246	BY SIMILARITY.	255	QY	602	QYIKANSKFIG----ITELPDEGACQCPINCTHSCVDLDDKGGPAEORA-----SPLTS	553
FT	DISULFID	259	BY SIMILARITY.	286	Db	591	HGV-----LGAKGPIYKYPDVONECRFCHENCTOGCKGPELQDCLGQTLVIGKTHLTM	644
FT	DISULFID	290	BY SIMILARITY.	286	QY	654	IVSAVVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOETELVELTEPLTPSGAMPNQAM	712
FT	DISULFID	305	BY SIMILARITY.	301	Db	645	ALTVIAG--LVVIFMMLGGTFLYWRGRIQNKAMRRVLERGESIEPLDPS-EKANKVLA	701
FT	DISULFID	320	BY SIMILARITY.	327	QY	713	RILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAY	772
FT	DISULFID	327	BY SIMILARITY.	327	Db	702	RIFKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAY	772
FT	DISULFID	330	BY SIMILARITY.	327	QY	773	VMAGVSPYVSRLGICLTSTVQLVTQMLPYGCLLDHVRENKRGRLGSGDOLLNMCQIAKG	832
FT	DISULFID	341	BY SIMILARITY.	327	Db	762	AIGSLDHAHIVRLGLGCPGSSQLVTVPLGSLLDHVRQHRGALGPGQLLNMGVOIAKG	821
FT	DISULFID	353	BY SIMILARITY.	353	QY	833	MSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALE	892
FT	DISULFID	408	BY SIMILARITY.	408	Db	822	MYLLEHGMVHRNLAAARNVLLKSPQVQVADFGVADLLPPDDKQLLYSEAKTPIKMALE	881
FT	DISULFID	414	BY SIMILARITY.	414	QY	893	SILRRRTHOSDVMWSYGVVWELMTFKAQKPYDGIIPAREIPOLLEKGERLPPICTIDVY	952
FT	DISULFID	437	BY SIMILARITY.	437	Db	882	SIHFQKTHOSDVMWSYGVVWELMTFKAQKPYDGIIPAREIPOLLEKGERLPPICTIDVY	941
FT	DISULFID	469	BY SIMILARITY.	469	QY	953	MIWVKWIMIDSECRPRELSEFSEMRARDPQRFVIONEDLGA-----SPLDSTFYRSL	1009
FT	DISULFID	522	BY SIMILARITY.	522	Db	942	MMVKWIMIDSECRPRELSEFSEMRARDPQRFVIONEDLGA-----SPLDSTFYRSL	1009
FT	DISULFID	566	BY SIMILARITY.	566	QY	1010	EDDMGDLVDAEYLPQOGFCFDPAPAGAGGMVHHRHSSTSGGDLTLGLEP-SEE	1068
FT	DISULFID	616	BY SIMILARITY.	616	Db	1001	EEVELEPELDLDLEAED-----NLATTLGASLSLPGTLNRP	1041
FT	DISULFID	141	BY SIMILARITY.	183	QY	1069	EAPRSLAPSEGAGSDVFDGLGMAAKGLOSLPTH-PSPLQRYSEDTPVPLP-----	1121
FT	DISULFID	184	BY SIMILARITY.	184	Db	1042	RGQSLLSPSSGY-MPMNQNLGESCQESAVSGSSERCPVSLH-----PMRGLAS	1094
FT	DISULFID	184	BY SIMILARITY.	184	QY	1122	SETDGYVA-----PLTCSQPE-----YVNOPVVRPOPPSPREP-----	1156
FT	DISULFID	184	BY SIMILARITY.	184	Db	1095	ESSEGHVTGSEAELOEKVSMCRSRSRSPRPGDSAYHSQHSLSLTPVTLSPPLGLEE	1154
FT	DISULFID	184	BY SIMILARITY.	184	QY	1157	-----LPAARPAGATLERAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGA	1203
FT	DISULFID	184	BY SIMILARITY.	184	Db	1155	DVNGYVMPDTHLKGTPSSREGTLSSVGLSSVLGTEDEED-----EYFYNNRRRH	1206
FT	DISULFID	184	BY SIMILARITY.	184	QY	1204	APQHPHPPAFSPAFDNLVYD-----QDPPERGAPSTFKGTPTAENPEYL	1249
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	QY	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID	184	BY SIMILARITY.	184	Db	1207	SP-PHPPRPSSUELEGYEMDVGSDLSASIGSTQSCPLHPVIMPTAGTTTDPDEYEM	1263
FT	DISULFID</							

RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96096535; PubMed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
RT "Cloning of the rat ErbB3 cDNA and characterization of the  
recombinant protein.";  
RN Gene 165:279-284(1995).  
RL (2)  
RP REVISIONS TO 85: 513 AND 565.  
RA Hellyer N.J., Koland J.G.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN (3)  
RP SEQUENCE OF 922-1097 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuroligins and their putative receptors, ErbB2 and  
ErbB3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659(1997).  
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
(POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
PHOSPHATIDYLINOSITOL 3-KINASE.  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U29339; AAC28498.2; -  
DR EMBL; U52530; AAC53050.1; -  
DR HSSP; P11362; IFGK.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR ProDom; PD01030; Recep\_L domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR SMART; SM000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 19  
FT CHAIN 20 1339  
FT DOMAIN 20 643 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.  
FT TRANSMEM 644 662 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 663 1339 POTENTIAL.  
FT DOMAIN 183 259 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 707 964 CY5-RICH.  
FT NP\_BIND 713 721 PROTEIN KINASE.  
FT BINDING 740 740 ATP (BY SIMILARITY).  
FT ACT\_SITE 832 832 ATP (BY SIMILARITY).  
FT DISULFID 186 194 BY SIMILARITY.  
FT DISULFID 190 202 BY SIMILARITY.  
FT DISULFID 210 218 BY SIMILARITY.  
FT DISULFID 214 226 BY SIMILARITY.  
FT DISULFID 227 235 BY SIMILARITY.

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SQ	SEQUENCE	1339	AA; 147545	MM; 0AASF2402BBFDF1E CRC64;
Query Match 34.2%; Score 2327.5; DB 1; Length 1339;				
Best Local Similarity 40.7%; Pred. No. 9.2e-118;				
Matches 524; Conservative 169; Mismatches 429; Indels 167; Gaps 36;				
QY	3	LAALCRWGLLALLPPGAA	---STQCTGTDMLRLPASPTHLDMLRLHYOGCVOQCN	59
DB	7	LOVLC	----FLUSLARGSEMNSQAVCPGTUNGSLVTDGADNQYOTLYKLYEKCEVVMGN	62
QY	60	LELTYPNTASLSFLQDIQEQGYVLI	AHNOVRQVPLQRRLRIVRGTTQFEDNYALAVLDN	119
DB	63	LEIVLTGHNADLSFLQWIREVTGYV	LNVAMNEFSLPLNLRVVRGTQVYDKFAIFVM--	120
QY	120	GDPLNNTPTVGTASPGGLREQLRL	TEILKGGVLIQRNPQLCYOQTLWKDOI	179
DB	121	---LNYNT	----NSSHALRQLKFTOLTILSGGVYIEKNDKLC	170
QY	180	ALTLDITNRACHPCSPMGKSRG	BESSDDCSLTTRVCAGGC-ARCKGPLPTDCCHE	238
DB	171	GAEIVVKNNGANGCPCPEVCVK	G-RCWGGPDDCQILTITICAPQCNGRCFGPNPQCCHD	229
QY	239	OCAAGCTGPKHSDCLACLFHNSG	ICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA	298
DB	230	ECAGGCGSPQDTDCFACRRFND	SGACVPCRCPEPLVYNKLTFOLEPNPHTKYQGGVCVAS	289
QY	299	CPYNYLSTDVGSCTLVCPHMQE	VTAEDGTORCEKSKPCARVCYGL--GMQYIKANSKF	356
DB	290	CPHNEV-VDQTFVCRACPPDKME	VD-KHGLKMCPCGGGLCPKACEGTSGSRYQTVDSSN	347
QY	357	IGITELEFAGCKKIFGSLAFPE	SDGDPASNTAPLOEQLOVFTLEITCYLISAWP	416
DB	348	ID----	GFVNCITKILGNLDFLITGLNVDFMWHKIPALDPEKLVFRTVREITGLNIQSWP	403
QY	417	DSLPLDSVFQNLQVIRGIRILH	NGAYS-LTLOGLGISWLGRLSLRLGLSGLAIIHNHTLIC	475
DB	404	PHMNFVSFSLNTITIGRSLN	RGFSLIMKNLNTVSLGFRSLKLSISAGRVVISANQQLC	463
QY	476	FVHTVPDQLPFRPHQALLHTA	-NRPEDECVGEGLACHOLCARGHCWGPPTQCVNCSOF	534
DB	464	YHSLNWTLLRGPSEERLDIKY	DRPLGELAEGRVCDPLCSSGGGCGWGPQGLSCRNY	523
QY	535	LRGQECVEECRVLOGLPREY	NARHCLCHPECQPNQSVTCFGEAOCVACAHYKQBP	594
DB	524	SREGVCVTHCNFLOQEPREF	VHEAQCFCSCHEPCLPMEGTSTCNGSGSDACARCAHFRDGP	583
QY	595	FCVARCPQYIKANSKFIG	----ITELPDEGACQPCPINCTHSC--VLLDDKGCPEAQR	648





```
Db 603 IADCGVISNAYK--FDNRCTCKHPCER-----TCNGAGADHCQCVHVRDQHCQVSC 654
Qy 601 POYIKANSK-----FIGITEL----- 616
Db 655 PKN-KYNDGVCRECHATCGCTGPKDTIGGACTTCNLAIINNDATVRRCLLKDDKCPD 713
Qy 617 -----PDEGACQP-----CPI-----NCHT----- 632
Db 714 GYFWEYVHPQEGSLKPLAGRAVCRKCHPLCELCNTNYGHEQVCCKTHYKRREOCETEC 773
Qy 633 -----SC-----VDLDKG-----C 642
Db 774 PADHYTDEORECFORHPECNGCTGPGADCKSCRNFKLFDANETGYPVNSTMFCNTSKC 833
Qy 643 PAEOR-----ASPLTS-----IVSAVVGILLVVLGVVFGILI 675
Db 834 PLEMRHVNYQYTAIGPYCAASPRSKYTANLDVNMIFITGAVLVPTICILCVV--TVI 891
Qy 676 KRQOKIRKYT--MRRLLQETELVEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGT 733
Db 892 CRQOKAKKETVMTMALSGCEDSEPLRPSNIGANLKLIRVQKDAELRKGVLGAGFR 951
Qy 734 YKGIWIPDGENVKIPVAKVLRNTSPKANKEILDEAYVMAGVSPVVSRLIGICLTST 793
Db 952 YKGVVWPEGENVKIPVAKELKSTGAESSEEFREAYINASEEHVNLKLLAVCMSSQ 1011
Qy 794 VOLVTOLMPYGLLDHVRNRLGSDLLNKCWQIAKMGMSYLEDVRLVHRDLAARNVLV 853
Db 1012 MMLITQLMPLGLLDYVNRNNDKIGSKALLNMSTQIAKMGMSYLEEKLRLVHRDLAARNVLV 1071
Qy 854 KSPNHVKITDFGLARLLDIDETEHADGCKVPKPKMALESILRRFTHQSDVWSGVTVW 913
Db 1072 QTPSLVKITDFGLAKLLSSDNEYKAGGKPKPKWALECIRNVRTSKSDVWAFGVTVI 1131
Qy 914 ELMTFGAKPYDGIPIAREIPDLEKGERLPQPICTIDVYIMVRCWMIDSECRPRFREL 973
Db 1132 ELLTFGQPHENIPAKDIPDLIEVLGKLEQPEICSDIYCTLLSCWHLDAAMRFTFKLT 1191
Qy 974 SEFSRMARDPQRFVVIQNEGLG--PASPLDSTFFRSILLEDD--DMGDLVDAAEYLVPPQ 1028
Db 1192 TVFAEFARDGRLYLAIPGDKFTRLP-----YTSQDEKDLIRKLAFTTDSIAIKPD 1244
Qy 1029 GFPCPDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEBAP-----RSPLAPSGAG 1082
Db 1245 DYLOPKAAPGFS-----HRTDCT-----DEMPKLNRYCKDPSNWSSTG 1283
Qy 1083 SDVFDG---DLGMAAGKLSQSLPHDPSPLQRYSEDTVPLPSETDGYVAPLTCSPQPEY 1139
Db 1284 DDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCQPGNN 1321
Qy 1140 VNQPDVVRPQPPREGPLPAARPAAGATLERAKTILSPGKNGVVKVDVAFGGAIVENPEYL-- 1197
Db 1322 NNNNN-----NPNQNNMAVGVAAGYM-----DLIGVPVSDVNPEYLLN 1360
Qy 1198 --TPQGAAPQPH-----PPAFSP-AFDNLYWD 1224
Db 1361 AQLGVGESPIQTIGIPWVGPGTMEVKVNPVPGSEPTSSDHEYND 1408
```

## RESULT 12

```
ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.,
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSRP; P11362; LFQK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase.1
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR F. Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 25.3%; Score 1719.5; DB 1; Length 634;
Best Local Similarity 51.8%; Pred. No. 1.8e-85;
Matches 366; Conservative 79; Mismatches 139; Indels 123; Gaps 17;
QY 587 CAHYKDPFCVRCPOYIKANSKFIGITELPDEGACQPCIPNCTHSCVDLDDKGPAAEQ 646
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTRCCKGPGLESCP--- 58
QY 647 RASPLTSIVSAVY-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGA 705
Db 59 NGSKTFSIAAGVVGGLLCLVVGIGIGLYLRRR-HIVRKETRLRLLOERELVEPLTPSGE 117
QY 706 MPNQAMRILKETELRKVKVLGSGAFGTGVKGIWIPDGENVKIPVAKVLRNTSPKANK 765
Db 118 APNQAHLRLKETEPKPKVKVLGSGAGFTVYKGLWIFEKVKIPVAKLEATSPKANK 177
QY 766 EILDEAYVMAGVSPVVSRLIGICLTSTVOLVQLMPYGLLDHVRNRLGSLQDILNN 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVOLVQLMPYGLLDVIREHKDNIGSQYLNN 237
QY 826 CMQIAKMGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGCKVP 885
Db 238 CVQIAGMNYLEERLVLHRDLAARNVLVKTPOHKITDFGLAKLLGADEKEYHAEGKVP 297
QY 886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLEKGERLPQPP 945
```



```
Db 298 IKWMALESILHRIYTHQSDVMSGYTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYIMVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSKWARDPPRYLVIQDERMHLPSPTDSKF 417
Qy 1005 YRSLEDDMDGLVDVAEYLVPOQGFPCDPAPGAGGVVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEAAPSPL-----APSEGAGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSDPTVP 1119
Db 450 -----SRTPLLSLSATSNNATCID-----RNGQHPVREDSFVQRYSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGK 1177
Db 496 FLEESIDGFL-----PAPEYVNO--LMPKKPS-----TAMVQ 526
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPAPAFSPAFDNLVY 1222
Db 527 NOIYNNISLTAISKLPMDSRYSQNSHSTAVDNPYL-----NTNQSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPPSTFKGTPTAENPEYLGLDVP 1254
Db 579 WIQSGNHQINLNDPDYQDQFLPNETKPNGLLKVAENPEYLRVAAP 625
```

## RESULT 13

```
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78 (1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459 (1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVIUI.
DR HSRP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132..399 PROTEIN KINASE.
FT NP_BIND 138..146 ATP (BY SIMILARITY).
FT BINDING 165..165 ATP (BY SIMILARITY).
FT ACT_SITE 257..257 BY SIMILARITY.
FT CONFLICT 29..29 R -> W (IN REF. 2).
FT CONFLICT 140..140 S -> F (IN REF. 2).
FT CONFLICT 146..146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BBD006745D0609 CRC64;

Query Match 24.8%; Score 1688; DB 1; Length 604;
Best Local Similarity 51.7%; Pred. No. 8.6e-84;
Matches 357; Conservative 76; Mismatches 131; Indels 126; Gaps 16;

Qy 587 CAHYKDPFCVARGPQYIKANSKFITELPDEGACQPCINCHSCVDLDDKCPABEQ 646
Db 3 CAHFTDGHPCVKACPAVLGENDTL-VKRYADANAVCOLCHPNCRTGCKGPGLEGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSCA 705
Db 59 NGSKTPTAAGVWGGLCLVVLGIGLYLRRR-HIVRKRTLRLRLQRELVEPLTPSGE 117
Qy 706 MPNOAQRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSFKANK 765
Db 118 APNOAHLRLKETEFKVKVLGSGAFGIYKGLWIPEGEKVKIPVAIKELRATSPKANK 177
Qy 766 EILDEAYVWAGVGSYVSRLLIGLICLTSTVQLVTQLMYPYGLLDHVRNRRIGSODLLNW 825
Db 178 EILDEAYVWASVDNPHVCRLLIGLICLTSTVQLITQLMYPYGLLDYIREHKDNIGSQVLLNW 237
Qy 826 CMOAKGMSYLEDVLRVHRDLAARNVLKSNHVKITDGLARLLDIDETEYHAGGKVP 885
Db 238 CVQAKGNVLEERLVRDLAARNVLKTPQHVKITDGLAKLLGADEKEYHAEKGKVP 297
Qy 886 IKWMALESILRRRFTHQSDVMSGYTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSGYTVWELMTFGSKPYDGIIPASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYIMVKCMWIDSECRPRFRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCMWIDADSRPKFRELIAEFSKWARDPPRYLVIQDERMHLPSPTDSKF 417
Qy 1005 YRSLEDDMDGLVDVAEYLVPOQGFPCDPAPGAGGVVHRRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEAAPSPL-----APSEGAGSDVFDGDLGMAAGKGLQSLTPHDPSPLQRYSDPTVP 1119
Db 450 -----SRTPLLSLSATSNNATCID-----RNGQHPVREDSFVQRYSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPGK 1177
Db 496 FLEESIDGFL-----PAPEYVNO--LMPKKPS----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPAPAFSPAFDNLVY 1222
Db 527 NOIYNNISLTAISKLPMDSRYSQNSHSTAVDNPYL-----NTNQSPLAKTVF 574
Qy 1219 NLYTWYDQDPPERGAPPSTFKGTPTAENPEY 1248
```



```

Db 575 SGPYIQQNHQ-----INLNDPDY 594

RESULT 14
ERBB AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein, erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103896;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RL cells."
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M13179; AAA2401.1; -
CC PIR: A25231; TVFVEB.
CC HSP; P11362; IFGK.
CC InterPro: IPR000719; Euk_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00219; TyKc; 1.
CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE: PS50011; PROTEIN KINASE DOM; 1.
CC Transferrase: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399 PROTEIN KINASE.
CC NP_BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT_SITE 257 257 BY SIMILARITY.
CC FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
CC SQ SEQUENCE 540 AA; 60412 MW; 5853297AA06865D CRC64;

Query Match
Best Local Similarity 54.4%; Pred. No. 4.1e-79;
Matches 337; Conservative 69; Mismatches 123; Indels 90; Gaps 14;

QY 587 CAHYKDPFPCVACRPOYIKANSKFIGITELPDEGACQPCPINCSTHSCVDLDDKGCPSAQ 646
Db 3 CAHFDGPHCVKACPAVGLGENDTL-VKPYADANAVCQLCHPCTRGCKGPGLEGCP--- 58
QY 647 RASPLTSIVSAVV-GILLVVVLGVFGILIKRQOKIRKYTWRRLLQETELVEPTPSGA 705
Db 59 NGSKTPSTAAGVGGULLCVVGLGIGLYLRR-HIVKRTLRLLQERLVEPTPSGE 117
QY 706 MPNOAMRILKTELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANK 765
Db 118 APNOAHLRLKETEFKVKVGLGFGAGTVYKGLWIPGEKVTPVAIKELRATSPKANK 177
QY 766 EILDEAYVMAGVGSYVSRLLGICLTSTVQLTQMPYGCCLLDVHRENRGLGSDQLLNW 825

```

```

Db 178 EILDEAYVMASVDNPHVCRLLGLCLTSTVQLITQLMPYGCCLLDVIREHKDNIGSQVLLNW 237
QY 826 CMQIAKMSYLEDVRLVHRDLAARNVLKSPHVKITDFGLARLLDIDETEHADGKVP 885
Db 238 CVQIAKGMVYLEERHWHVHRDLAARNVLVKTPOHVKITDFGLAKQLGADKEKEYHAEGKVP 297
QY 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPPOP 357
QY 946 ICTIDVYIMVKCWMIDSECRFRRELVSFMRMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYIMVKCWMDSADSRKRELIAEFKMDARDPPYLVIVIQGERMHLPSPTDSKF 417
QY 1005 YRSLLDEDDMGDLVDAEEVLPQGFPCDPAPAGGVMVHHRSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 449
QY 1065 PSEEAAPRSL-----APSEGAGSDVFDGDLGMAAGKGLQSLPTHDPSPLOQRYSEDP 1119
Db 450 -----SRTPLLSLSLATSNSATNCIDRNG-----H----- 476
QY 1120 LPSETDGVVAPLTCSQPEYVNPQDVRPQPPSPREGPLPAARPAAGAT-LERAKTLSPGKN 1178
Db 477 -PVREDGFL-----PAPEYVQ--LMPKKEPSTAMVQNIYISLTAISKLPIDSRVQN 527
QY 1179 GWKXDVFAFGGAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR CHICK STANDARD; PRT; 703 AA.
ID EGFR CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
DE EGFR.
GN Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```

```
or send an email to license@iab-sib.ch).
-----
CC      EMBL; M20386; AAA48760.1; -. domain.
DR      InterPro; IPR000494; EGFR_L_domain.
DR      InterPro; IPR000719; Euk_pkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      SMART; SM00261; FU; 4.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR      PROSITE; PS00111; PROTEIN_KINASE_DOM; PARTIAL.
KW      Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW      Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT      SIGNAL      1..30
FT      CHAIN      31..703
FT      DOMAIN      31..654
FT      DOMAIN      31..654
FT      TRANSMEM      655..667
FT      DOMAIN      668..703
FT      DISULFID      197..206
FT      DISULFID      201..214
FT      DISULFID      222..230
FT      DISULFID      226..238
FT      DISULFID      239..247
FT      DISULFID      243..255
FT      DISULFID      258..267
FT      DISULFID      271..298
FT      DISULFID      302..314
FT      DISULFID      318..333
FT      DISULFID      336..340
FT      DISULFID      513..522
FT      DISULFID      517..530
FT      DISULFID      533..542
FT      DISULFID      546..562
FT      DISULFID      565..581
FT      DISULFID      569..589
FT      DISULFID      592..601
FT      DISULFID      605..627
FT      DISULFID      630..638
FT      DISULFID      634..646
FT      CARBOHYD      134..134
FT      CARBOHYD      190..190
FT      CARBOHYD      200..200
FT      CARBOHYD      359..359
FT      CARBOHYD      368..368
FT      CARBOHYD      420..420
FT      CARBOHYD      573..573
FT      CARBOHYD      578..578
FT      CARBOHYD      613..613
FT      CARBOHYD      633..633
FT      CARBOHYD      648..648
FT      NON_TER      703
SQ      SEQUENCE      703 AA; 77427 MW; AFF2DE11B735A690 CRC64;

Query Match      23.0%; Score 1565; DB 1; Length 703;
Best Local Similarity      44.1%; Pred. No. 4.2e-77;
Matches 312; Conservative 111; Mismatches 255; Indels 30; Gaps 14;

Qy      8  RWGULLALLPPGAA-----STQVCTGTDMLKRLPASPETHDMLRHLHYQCQVVGNNLE 61
Db      13  RGAALVLLLLGVALCSAVEEKVCQGTNNKLTQLGHVEDHFTSLQRMYNNECVVLSNLE 72

Qy      62  LTYLPTNATSLFLODIQGVQVLIAHNVQVPLQRLRIVRGTLQFEDNYALAVLNGD 121
Db      73  ITTYVHNRLDTFLTKIQVAGVGLVIALNMVDVPLENLIQIRGNVLYDNSFALAVLSNYH 132

Qy      122  PLNNTTPTVTGASPGGLRELQRLSLTEILKGGVLIQVLPOLCYQDTILWKDIFHKNNQLAL 181
Db      133  -MNKTQ-----GLRELPMKRLSEILNGGVKLSNNPKLCNMDTVLWNDIIDTSRK-PL 182

Qy      182  TLID-THRSRACHPCSPWCKGSRGWGSSSDCQSLTRTVACAGGCA-RCKGPLPTDCCHEQ 239
Db      182  TLID-THRSRACHPCSPWCKGSRGWGSSSDCQSLTRTVACAGGCA-RCKGPLPTDCCHEQ 239
```

```
Db      183  TVLDFASNLSSCPKCHPNCTEDHCWAGEQNCQTLTKVICAQCSGRCKGVKVPSPDCCHNQ 242
Qy      240  CAAGCTGPKGSHDCLACHFNHSGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTAC 299
Db      243  CAAGCTGPRSDCLACRKFRRDADATCKDTCPPLVLYNPTTYQMDVNPPEGKYSFGATCVREC 302
Qy      300  PYNVLSLTDVSGCTLVCPLNQEVTAEDGTORCEKSKPCARVCVGLGMOYIKANSKFYGI 359
Db      303  PHNVVTDHSGSVRSCTNTDTYEV--EENGVRKCKKCDGLCKVCGNGIGIGELKGLS-INA 360
Qy      360  TELE-FAGCKKIFGSLAPFSPDFDGPASNTAPLOEQLOVFETLEETITGVLYISAWPDS 418
Db      361  TNIDSFKNCTKINGDVSLPVAFLGDAFTKTLPLDPKKLDVFRVTVKEISGFLLIQAWPDN 420
Qy      419  LPDLSVFQNLQVIRGRILHNCAYSILTQGLGISWLGSLRELSGSLALIHNNHLCFVH 478
Db      421  ATDLYAFENLEIIRGRTKQHGQYSLAVVNLKIQSLGLSLKEISDGDIAIMKNKLCYAD 480
Qy      479  TVPMDQLFRNPHQALLHTANRPEDECVEGEGLAHQLCARGHCWGPGTQCVNCSOFLRGO 538
Db      481  TMNWSLFAFOSQKTIIQNRKNDCTADRHVCDPLCSDVCGWGPFPCHFCSEFFSRQK 540
Qy      539  ECVBEERVLOGLPREYVNAHCLPCHPECPQNG---SVTCFGEADQCACAHYKDPFF 595
Db      541  ECVKQCNILOQEPREFERSKCLPCHSECLVQNSTAYNTTCGPGPDHCKCAHFIDGPH 600
Qy      596  CVARCPQVIKANSKFIGITELPDEGACQPCPINCTHSCVDLDDKGCBAEQRASPLTSIV 655
Db      601  CVKACPAGVLGENDTL-VMKYADANAVCQLCHPNCTRCKGKGLGEGCP---NGSKTPSIA 656
Qy      656  SAVV-GILLVVVLGVVFGILIKRQOKIRKYVTMRLLQETELVEPLTP 702
Db      657  AGVVGGLLCLVVVGLGIGLYLURR-HIVRKRLRLRLQERELVEPLTP 703
```

Search completed: July 22, 2003, 08:46:51

Job time : 21.2304 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds  
(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-465-479-12  
Perfect score: 6804  
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPSYGLGIDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6129	90.1	1259	6	O18735
2	3114	45.8	1209	11	Q9GX70
3	3085	45.3	1210	11	Q9EP98
4	2716	39.9	1165	13	Q9YH40
5	2667.5	39.2	1137	13	Q9W6F6
6	2289	33.6	1328	13	P79754
7	2007.5	29.5	1433	5	Q9BIH9
8	1871	27.5	419	4	Q9UK79
9	1739	25.6	367	11	Q8R2X1
10	1720	25.3	729	15	Q86712
11	1718	25.2	567	15	Q86714
12	1697.5	24.9	412	4	Q8WYV0
13	1653.5	24.3	962	15	Q64895
14	1645	24.2	545	15	Q85468
15	1476.5	21.7	655	11	Q9WVF5
16	1460.5	21.5	643	11	Q9ERV6

17	1274	18.7	1193	5	Q9Y1X8	Q9Y1X8 ephydatia f
18	1182.5	17.4	1368	5	Q23821	Q23821 caenorhabdi
19	1169	17.2	1717	5	Q26566	Q26566 schistosoma
20	1108	16.3	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9SE80	Q9SE80 rattus norv
22	928.5	13.6	599	13	Q9PSH2	Q9PSH2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	806.5	11.9	346	13	P11776	P11776 xiphophorus
25	800	11.8	176	11	O923V5	O923V5 rattus norv
26	778	11.4	435	5	O8SZW1	O8SZW1 drosophila
27	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
28	734	10.8	331	4	Q9BUD7	Q9BUD7 homo sapien
29	731.5	10.8	1362	13	Q9PVZ4	Q9PVZ4 xenopus lae
30	723	10.6	149	6	Q9BG66	Q9BG66 oryctolagus
31	721	10.6	1671	5	Q9NJV5	Q9NJV5 biomphalari
32	695	10.2	1418	13	O93457	O93457 scophthalmu
33	685.5	10.1	1368	13	Q8UW85	Q8UW85 paralichthy
34	672.5	9.9	1369	13	Q8UW86	Q8UW86 paralichthy
35	664.5	9.8	1472	5	Q9U5A8	Q9U5A8 bombyx mori
36	660	9.7	1358	13	O73798	O73798 xenopus lae
37	658	9.7	1412	13	Q8UW84	Q8UW84 paralichthy
38	641.5	9.4	1418	13	Q8UW83	Q8UW83 paralichthy
39	638	9.4	1245	13	Q9YGH8	Q9YGH8 scophthalmu
40	628	9.2	1371	11	Q9QVW4	Q9QVW4 rattus sp.
41	620.5	9.1	2144	5	Q9VD94	Q9VD94 drosophila
42	600	8.8	987	11	Q91YMO	Q91YMO mus musculu
43	595	8.7	987	11	Q99MR2	Q99MR2 mus musculu
44	593	8.7	935	4	Q96L35	Q96L35 homo sapien
45	587.5	8.6	1036	4	Q07912	Q07912 homo sapien

#### ALIGNMENTS

#### RESULT 1

O18735 ID O18735 PRELIMINARY; PRT; 1259 AA.  
AC O18735;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Erbb-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erbb-2 from canine mammary gland."  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA2127.1; -  
DR HSSP; P11362; IFGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP\_2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM0261; FU; 3.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		90.1%; Score 6129; DB 6; Length 1259;
Best Local Similarity		90.0%; Pred. No. 0;
Matches 1134; Conservative		42; Mismatches 78; Indels 6; Gaps 2;
QY	1	MELAAALCRWGLLALLPPGAASFOVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60
DB	1	MELAAWCRWGLLALLPSGAAGFOVCTGTDMLKRLPASPTHLDMLRHLYQGCVVQGNL 60
QY	61	ELTYLTNASLSFLODIQEVQGVVLIHNOVROVPLQRLIRIVRGTLQFDNYALAVLDNG 120
DB	61	ELTYLPANASLSFLODIQEVQGVVLIHNSQVROIPLQRLIRIVRGTLQFDNYALAVLDNG 120
QY	121	DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCVQDTILWKDIFPHKNQLA 180
DB	121	DPLEGGIPAPGAAGGLRELQRLSLTEILKGGVLIQPSPOLCHQDTILWKDVFHKNQLA 180
QY	181	LTLIDTNRSRACHPCSPCKSGCWSESSEDCSLTRTVCCAGCARCKGPLPTDCCHEOC 240
DB	181	LTLIDTNRFSACPSPCKADAHWCWSSGDCOSLRTTVCCAGCARCKGPPQPTDCCHEOC 240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300
DB	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTSCP 300
QY	301	YNYLSTDVSGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGMQVYKANSKFIGIT 360
DB	301	YNYLSTDVSGSCTLVCPHNLQEVTAEDGTQCEKSKPCARVCYGLGWEHLREVRVTSAN 360
QY	361	ELEFACKKIFGSLAFPSFDGPASNTAPLOEOLVFEETLEETIGLYISAWPDSLP 420
DB	361	IOEFACKKIFGSLAFPSFDGPASNTAPLOEOLVFEETLEETIGLYISAWPDSLP 420
QY	421	DLVSFONLQVIRGRIHNGAYSLLTQGLGISWGLSLRELGSGLALIHNNHLCFVHTV 480
DB	421	NLSVFONLQVIRGRVLDHGNYSLLTQGLGISWGLSLRELGSGLALIHNNARLCFVHTV 480
QY	481	PMDQLPROVYKANSKFIGITELECEVGEGLACHOLCARGHWGPGPTQCVNCSOFLRGQEC 540
DB	481	PMDQLFRNPHQALLHSANRPEEBECVGEGLACYP-CAHGHWCWGPQTQCVNCSOFLRGQEC 539
QY	541	VVECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600
DB	540	VVECRVLQGLPREYVKDRVCLPCHSECOPONGSVTCFGEADOCVACAHYKDPFPCVARC 599
QY	601	PSGKVPDLSPYMPKPPDEGACQPCPINCTHSCVDLDDKCPAEQASPLTSIVSAVVG 660
DB	600	PSGKVPDLSPMPKPPDEEGTCQPCPINCTHSCADLDEKCPAEQASPLTSIIAAVVG 659
QY	661	ILLVVVLGVVFGILLIKRQOKIRKYTMRLLOQTELVEPLTPSGAMPNQAQMRILKETEL 720
DB	660	ILLAVVVLVLGILLIKRRQKIRKYTMRLLOQTELVEPLTPSGAMPNQAQMRILKETEL 719
QY	721	RKVVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTSPKANKILDEAYVMAGVGSPP 780
DB	720	RKVVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTSPKANKILDEAYVMAGVGSPP 779
QY	781	YVSRLLGICLTSTVQLVTLMPYGCILLDHVRENRGLSGQDNLNWCQIAKMSYLEDYR 840
DB	780	YVSRLLGICLTSTVQLVTLMPYGCILLDHVREHRLSGQDNLNWCQIAKMSYLEDYR 839
QY	841	LVRDLAARNVLKSPNNHKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRTT 900
DB	840	LVRDLAARNVLKSPNNHKITDFGLARLLDIDETEHADGGKVPKWMALLESIPRRRTT 899
QY	901	HOSDVMYSYVTVWELMTFCAKPDVGIIPAREIPDLLEKGERLPQPICTIDVYMWKCM 960
DB	900	HOSDVMYSYVTVWELMTFCAKPDVGIIPAREIPDLLEKGERLPQPICTIDVYMWKCM 959
QY	961	IDSECRPRELVSFSESRMARDPQRFVJQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1020
DB	960	IDSECRPRELVAEFSRMARDPQRFVJQNEDLGPASPLDSTFYRSLLDEDDMGDLVDA 1019
QY	1021	EYLVPOQGFCCPDPAPAGGWHHRSSSTRSGGDLTLGLEPSEEEAPRPLAPSE 1080
DB	1020	EYLVPOQGFCCPDPAGGTAHRRSSSTRNGGDLTLGLEPSEEEPPKPLAPSE 1079
QY	1081	AGSDVFDGLGMAAGLQSLPDRYSEDPVPLPSETDGVVAPLTCSPQPEYV 1140
DB	1080	AGSDVFDGLGMAAGLQSLPDRYSEDPVPLPSETDGVVAPLTCSPQPEYV 1139
QY	1141	NQPDVPRQPSPREGPLPAAPAGATLER-----AKTSPGKGVVVKDVFAGGAVENPE 1195
DB	1140	NQPEVWPQPLALEGLPSPRPAATLERPKTSLSPGKGVVVKDVFAGGAVENPE 1199
QY	1196	YLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPSTFTKGTPTAENPEYGLDVPV 1255
DB	1200	YLAPRGAAAPQHPHPPAFSPADNLYYWDQDPSERGSPPSTFEGTPTAENPEYGLDVPV 1259
RESULT 2		
Q9QX70		PRELIMINARY; PRT; 1209 AA.
AC	Q9QX70;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OC	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RX	MEDLINE=90758888; PubMed=2342466;	
RA	Petch L.A., Harris J., Raymond V.W., Blaasband A.J., Lee D.C.,	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue."	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K., Dawson T.L., Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; .	
DR	HSSP; P11362; 1FGK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B7773 CRC64;	
Query Match		45.8%; Score 3114; DB 11; Length 1209;
Best Local Similarity		50.0%; Pred. No. 4.7e-226;
Matches 639; Conservative		165; Mismatches 358; Indels 116; Gaps 28;
QY	3	LAALCRWGLLALLPPGA-ASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNLE 61
DB	3	LAALCRWGLLALLPPGA-ASTQVCTGTDMLRLPASPTHLDMLRHLYQGCVVQGNLE 61





SQ	SEQUENCE	1165 AA; 129614 MW; 7F7E3B8D771A74E CRC64;
Query Match	39.9%; Score 2716; DB 13; Length 1165;	
Best Local Similarity	45.6%; Pred. No. 5.1e-196;	
Matches	582; Conservative 162; Mismatches 384; Indels 148; Gaps 31;	
QY	1 MELAAICRWGILLALLPG-AAST-----QVCTGTDMKRLRLPASPTHLDMLRLHLYQGQCV 55	
DB	4 LELLE-----LLLLLSIGRCSTDPDRKVCQGTSNQMTM-----LDNHYLKKMKMYSGCNV 56	
QY	56 VQGNLELYPTNASLFLQDIOEVGVVLIHNOVROVPLQRLIRVGTOLFEVDNYALA 115	
DB	57 VLENLEITYQENQDLSFLOSIOEVGVVLIHNEVSTIPLVNLRLIRGQNLVEGNFTLL 116	
QY	116 VLDNGDPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILWKDIFHK 175	
DB	117 VMSNYQK-NPSSP--DVYQVGLKQLQLNLTEILSGVKVSHNPLLCNVETINMWDIVDK 173	
QY	176 NNQALALTIDNRSACHPCSPMKSGSCWGESSEDCQSLRTVVCAGGC-ARCKGPLTD 234	
DB	174 TSNPTMNLIPHAERQCKQKQPCGVCNGSCWAPGPHGCHQKFTKLLCAEQCNRRRCRGPXPID 233	
QY	235 CCHECAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTPESMPNPEGRYTFGAS 294	
DB	234 CCNEHCAGCTGPRATDCLACRDFNDGTCKDTCPPKLYDIVSHQVVDNPNIKYTFGAA 293	
QY	295 CVTACPNYLTSDVGSCTLVCLPHNQEVTAEDGTQRCBCKSCPCARCVYGLGM-----QYI 350	
DB	294 CVKECPSYVYTE-GACVRSACSAGMLEVD-ENGKRSCKPCDGVCPKVCDDGIGLSNTI 351	
QY	351 KANSFIGITELEFAGCKKIGSLAFLESDGDPASNTAPLOPQLQVFTLEITGYL 410	
DB	352 AVNSTNIG-----SPSNCCTKINGDIILNRNSFEGDPHYKIGMPDPEHLNWLNTVTKBITGYL 407	
QY	411 YISAMPDLPDLSVFQNLQVIRGILHNGAYS-LTQGLGIGSWLGRSLRLSGSLALIH 469	
DB	408 VIMWPNWNTSLVSFQNLLEIRGRTTFSRGFSFVVVQVSHQWLGLSLKEVSAGNVIK 467	
QY	470 HNTLFCFVHTVPMQLOFROYIKANSKFIIGITELECVGEGGLACHQLCARGHCWGPPTQCV 529	
DB	468 NTPQLRVASTINWRLFRSEDSQ-----IEYDARTENQCNBESDGCWGPPTMVCV 520	
QY	530 NCSQFLRGQECVSCRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQVCAAH 589	
DB	521 SCLHVDGRGRCVASCNLLQGEPREAQVDGRVCQHQCELVQDLSLTCYGPFGPANCCKAH 580	
QY	590 YKDPFFCVARCPGSKVPDLSYMPIWKFPDEBACQPCPINCTHSCVDLDDKGCPEAQRAS 649	
DB	581 FQDGPQCIIPRCPHGMGLGDGDL-1WKYADKMGQCPQCHQNTQCGSGPGLSGCRGD-IVS 638	
QY	650 PLTSIVSVAVGILLVGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQ 709	
DB	639 HSSLAVGLVSGLLITIVALLIVVLLRRRIK-RKRTIRRLQELVEPLTPSGQADNQ 697	
QY	710 AOMRILKETELKVKVGLSGAGFTYKIGIWDGENVKIPVAIKVIRENTSPKANKEILD 769	
DB	698 AFLRILKETEPKORVLGSGAGFTYKGLWMPDGENIRIPVAIKVIREATSPKVNQEVLD 757	
QY	770 EAYVMAGSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLGSDLLNWCQI 829	
DB	758 EAYVMASVDHPVCRLLGICLTSAVLVTQMLPYGCLLDYVQHQERICGQWLLNWCQI 817	
QY	830 AKGMSYLEVRLVHRDLAARNVLKSPNHVKITDPLGLARLLDIDETEHADGKGVPIKWM 889	
DB	818 AKGMNYLEERLHVRLDLAARNVLLKNPNHVKITDPLGLSKLLTADKEYQOAHGKGVPIKWM 877	
QY	890 ALESILARRFTHQSDVMSYGVTVWELMTFGAKPDVGIIPAREIPDLLEGERLPOPPICTI 949	
DB	878 ALESILQWYTHQSDVMSYGVTVWELMTFGSKPDYDIPAKIASVLENGERLPOPPICTI 937	
QY	950 DVTMIMVKCWMIDSECRPRRELVSFRMRARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1009	
DB	938 EVYMIILKCMWIDFSSRPRRELVSFRMRARDPQRFVVIQNEIDLGPASPLDSTFYRSL 994	

QY	1010 EDDMDGLVDAAEYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLPSEBE 1069	
DB	995 SSDD--DVVDADAYLL-----RYKRIN-ROGS----- 1018	
QY	1070 APRSPLASEGAGSDVFDGLCMGAAGKLSLPHDPSPLQRYSEDPTV-PLPSETDGV 1128	
DB	1019 ---EPCIPNGH-----PVRENSIALRYISDPTQNALEKLDLGH- 1054	
QY	1129 APLTCSQPEYVNOVDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSP 1175	
DB	1055 -----EYVQPGSETSRSLSDIYNPNVEDLTDGMPVSLSSQEAETNFSREYLT 1105	
QY	1176 GKNGVVDVAFAGGAVENPEYLTPOGGAPOPHPPAFSPAFDNLYYMDQDPPERGAPS 1235	
DB	1106 NQNSL---PLVSSGSMDDPDY---QAG-----YQAF-----LPQTGALTG 1140	
QY	1236 TFKGTPTAENPEYLG 1251	
DB	1141 NGMFLPAAENLEYLGL 1156	
RESULT 5		
Q9W6F6	PRELIMINARY; PRT; 1137 AA.	
ID	Q9W6F6;	
AC	Q9W6F6;	
DT	01-NOV-1999 (TREMBLrel. 12, Created)	
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	Receptor tyrosine kinase (Fragment).	
GN	ERBB4.	
OS	Gallus gallus (Chicken).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_taxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=HINDBRAIN;	
RX	MEDLINE=99263203; PubMed=10328884;	
RA	Dixon M., Lumsden A.;	
RT	"Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain."	
RL	Mol. Cell. Neurosci. 13:237-258(1999).	
DR	EMBL; AF121963; AAD31764.1; -	
DR	HSSP; P11362; IFGK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001368; TNFR_c6.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	InterPro; IPR004019; YLP_motif.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L_domain; 1.	
DR	Pfam; PF02757; YLP; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrcK; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00111; PROTEIN KINASE DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.	
KW	Kinase; Tyrosine-protein kinase.	
FT	NON_TER 1	
SQ	SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;	
Query Match	39.2%; Score 2667.5; DB 13; Length 1137;	
Best Local Similarity	46.6%; Pred. No. 2.2e-192;	
Matches	534; Conservative 172; Mismatches 350; Indels 89; Gaps 28;	
QY	161 LCYQDTILWKDIFHKNQLALTLIDNRSACHPCSPMKSGRCWGESSEDCQSLRTVC 220	

```
Db 3 LCFADTIHWQDIVRNPWASNETLVPTNGSSGCRCHKSKCTG-RCWGPTEHNCQTLTKTVC 61
Qy 221 AGGC-ARCKGGLPTDCHEQCAAGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTF 279
Db 62 AEOCGRCYGYPVSCCHREAGGCGSPDIDCFACMFNDGACVTCQPTFVYVNPPTF 121
Qy 280 ESMNPENGRYTFGASCVTACPYNYLSTDVSGCTLVCPPLHNOEVTAEADGTORCEKSKPCA 339
Db 122 QLEHNHNKATYTGAFCKVKCPHNFV-VDSSSCVACRACPSKMEV-EENGIKMKCPCTDIP 179
Qy 340 RVCYGLGNOYIKANSKFIGITELE-FACKKIFGSLAFLPESFDGDPASNTAPLOPELOQ 398
Db 180 KACDGIQGSGL-VSAQTVDDSNIDKFINCTKINGNLIFLVTCIHGDYHTTAAINPEKLN 238
Qy 399 VFETLEETGYLYTSANPDSLPDLSVFONLQVIRGIRLHNGAYSILTQGLGISMGLRSL 458
Db 239 IFQTVREITGYLNTQSWENMTDFRVFNLVTIGRALYSGLSLLILKQOQITSLOFQSL 298
Qy 459 RELSGGLALIHNNTHLCFVHTVPMDOLF-----ROYIKANSKFIGITELECVGEGLACHQ 513
Db 299 KQISAGNIYITDNSNLCTYHTVNTSLFSTPSQKTVIHRNKK-----AENCTADGMVCNE 353
Qy 514 LCARGCHGPGTQCVCNSQFLRGQECVEBRLVLOGLPREYVNAHCLPCHPEQOP-QNG 572
Db 354 LCSSDGCNGPGPDQCLCKRIRGRTCTIESCNLYDGEFREFANGSVCMCEDPQCEKMDN 413
Qy 573 SVTCFPGPADOCVACAHYKDPFPCVARGPSGVKPDLSYMPITWKPDEBGACQPCINCH 632
Db 414 MITCYGPGPDHCTKCFHFKDGNPCVEKCPDGLQGNF--IFKYADEDEBCHPCHNCTQ 471
Qy 633 SCVDLDDKGC-----PAEORASPLTSIVSAVV-GILLVVVLGVVFGIILKRQQ 680
Db 472 GCRGPASHDCIYYPWTRQSTLPQHAR-TPL--IAAGVIGGLFIIVIMGLTFAVYVRKSI 528
Qy 681 KIRYVWERLLQETVELPELTPSGAMPNQAQWRILKETELRKVKVLGSGAGTYVKGWI 740
Db 529 K-KRAURRFL-ETVELPELTPSGAPNQAQRLILKETELRKVKVLGSGAGTYVKGWI 586
Qy 741 PDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVSPVSVRLGLICLSTVQLVTOL 800
Db 587 PEGETVKIPVAIKILNETTGPKANVEFNDEALIMASMDHPLHLLVGLVCLSPFTQLVTOL 646
Qy 801 MPYGLLDHVRNRRGLSGDQLNMCQIAKMSYLEDRVLVHRDLAARNLVKSPNHVK 860
Db 647 MPHGCLLDYVHEHKDNIQSLLNMCVQIAKGMVLEBRLLVHRDLAARNLVKSPNHVK 706
Qy 861 ITDFGLARLLDIDETEVHADGKVPKIMMALESILARRFTHQSDVMSYGVTVIWMELTGA 920
Db 707 ITDFGLARLLLEGDEKEYNADGKMPKIMMALECIHYRKFTHQSDVMSYGVTVIWMELTGG 766
Qy 921 KPYDGIIPAREIPDLLEKGERLPQPPICTIDYVMIMVKCMIIDSECRPRFRELVSFEFSMA 980
Db 767 KPYDGIIPAREIPDLLEKGERLPQPPICTIDYVMIMVKCMIIDASRPKFEKELAEFSMA 826
Qy 981 RDPQRFVVIQNEED-LGPASPLDSTFYRSLLEDODMDGLVDABEYLVPOQGFPCPDPA 1039
Db 827 RDPQRYLVIOGDDRMKLPSPNDSKFFQNLDEEDLEMDMADEYLVLP-QAFNIPIPIYTS 885
Qy 1040 GGMVHHRHSRSTSGGDLTLGLEPSEEAAPRS--PLAP-SEGASGVDFGDLGMGA 1096
Db 886 RTRIDSNRNQFVYRDGGVAEQGV-PMPYRAPGCIIPAPVAQATAEIPEDTCCNGFLR 944
Qy 1097 GLOSLTPHPSFLQRYSDPTVPLPS-----ETDGYVAPLTCSPQEPYVNPQDVRPOP 1149
Db 945 KOVATLAKEDSSTORYSADPTVFIPERVIRGELDEDDGYMTMRDKPKTDYLNPNVEENPFV 1004
Qy 1150 PSPREGPLPAA-RPAGATLERAKTSLPGKNGVVKDVF-----AFGAVENPEYLTPO 1200
Db 1005 SRRKNGDLQAVDNPYHN-----APNGQPKADEYVNEPLNTFANTLENAEYL--- 1054
Qy 1201 GGAAPQHPPPAFSPADNLVYDODPPERCA--PPSTFKCTPT-----AE 1244
```

```
Db 1055 -----KNNLPEKAKKAFDNDPDYWNHSLPPRSTLQHPDYLOEYSTKYFYKQNGRIRPIVAE 1109
Qy 1245 NPEYL 1249
Db 1110 NPEYL 1114

RESULT 6
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ErbB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OC NCBI_TaxID=31033;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 33.6%; Score 2289; DB 13; Length 1328;
Best Local Similarity 40.2%; Pred. No. 1.1e-163;
Matches 518; Conservative 157; Mismatches 420; Indels 192; Gaps 32;

Qy 9 WGLLLALLPP--GAASTQ---VCTGYDMKLRLPASPETHLMLRLHLYQGQVVOGNLEL 62
Db 4 WRLILMCVASRLRAASSTQAEVCPGTQNGLSSTGSQENQYNLNDKRYKGEIINGNLEI 63
Qy 63 TYLPTNASLSFLQDIEVQGVYVLIAHNQVROVPLQRLRIVRGTOQLFEDNYALAVLDNGDP 122
Db 64 TQIESNWDFFSLKTIREVTVGVLIAMNHQFPIPLQLRVRIGNSLYERRFALSVELN--- 120
Qy 123 LNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHNKQLALT 182
Db 121 ----YPKDG--PSGLNGLMLNLTEILDGGVQIINNLYRYPVWYRDII--RNNDAPIE 173
Qy 183 LIDNRSRACHPCSPMKGSRGSESSDQSLRTRTYCAGGC-ARCKGPLPTDCHEOCA 241
Db 174 IQFNGERGVCVH---KSC-GNYCWGPGKDKQQLITKTVCAPOCNDRCFTGSPRDCCHIECA 229
Qy 242 AGCTGPKHSDCLAHFNHSGICELHCPALVTYNTDTFESMNPENGRYTFGASCVTACPY 301
Db 230 AGCKGPLDTCFACRLFENDSGACVPQCPQTLIYNKQTFQMETNPNKAYQYSGICVSQCTP 289
Qy 302 NYLSTDVGSCTLVCPPLHNOEVTAEADGTOR-CEKSKPCARVCYGLGMOYIYKANSKFIGIT 360
Db 290 HFV--VDGSSCVSVCPDKMEV--ERGSQRCQELCSGLCPKYCEGTGAE-----ORQTVDS 342
```





```
QY 489 YIKANSKFIGITE-----LECVEGLACHQLCARGHMGPGTQCVCNCSQFLRGQECVBE 543
DB 446 -IKSSDHEVMQKRNATECHEEGECSEQCSKAGCWKGPEQCLECNWYKVKCLDS 504
QY 544 CRVLOGLPREY-VNARHCLPCHPECOPONGSVTCFGEADQCACAHYKDPFVCVACP- 601
DB 505 CK---SLPRLYSVDSKTCODCHQCKD-----FCYGNEDNCGSCMNKVGDRFCVABECP 556
QY 602 -----SGVKPDLSPVMPWKEPD--- 618
DB 557 TKHAMNGTCINCHKTCVGRGRDRTIADPGCISCDKAIIGSDAKTERCLMKDESPDGYV 616
QY 619 -----BEG----- 621
DB 617 SDYVLQEBGLKQLSGKAVCRKCHPRCKKCTGYGFHEQFCQECTGYKKGEQCEDECPDF 676
QY 622 -----ACQPCPINCT-----HSCVDL-----DD-----KGCPSAQ----- 646
DB 677 YANEETRICLPCHQECROCHGLGDHHECRNLKLFEGDPYDNATFTTCVSNCPASHPYKR 736
QY 647 -----RASPLTSIVSAVGVILLVVLVGVVFGI---LIKRRQKI 682
DB 737 PPOEAGKIGPYCSADMSQGLRIEPTQVKIVGSMVMAILLVCVFGIAFLVFSRHKKK 796
QY 693 RKYTHRRLLQETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAFGTYYKGIWIPD 742
DB 797 DAVKMTMALAGCEDSEPLRPSNVGNPLTKLRIKEAETRGGVLCMGAFGRVFKGVWMP 856
QY 743 GENVKIPVAIKVLRNTPSKANKEITLDEAYWAGVGSFYVSRLLGICITSTVOLUTQLMP 802
DB 857 GESVKIPVAIKVLMEMSESESEKFELEAYINASVEHPNLKLLAVCMSTQMLLTQLMP 916
QY 803 YGCLLDHVRNRRGLSGODLNCMCQIAKMSYLEDVRLVHRDLAARNVVKSPNHWKIT 862
DB 917 LGCLLDYRNKDKTIGSKALLNWSQIARGMAYLEERLVRDLAARNVLTQTPSCVKIT 976
QY 863 DFLARLLDIDETEHADGGKVPIIKWMALESILRRRFTHQSDVWSYGYVTVMELMTFGAKP 922
DB 977 VFLAKLLDFDSDEYRAAGGKMPIKWALECIRHVFVTSKSDVWAFGITWELLTYGAP 1036
QY 923 YDGPAREIPOLLEGERLPQPICTIDVYIMVWKWIDSECRFRRELVSFSEFMARD 982
DB 1037 YENVPAKDVPELIEIGHKLQPDICSLDYICLLSCWVLDADARPTFKQLAETFAEKARD 1096
QY 983 PORFVVIQNEDLGPASPLDSTFYRSLLDEDDMDLV----- 1018
DB 1097 PORYLM-----PGDKFMRLPSTYNQDEKDLIRTLAPVMAAAAAAAGASNVUV 1147
QY 1019 -----DABEYLVOQGFPCPDPAAGAGMVHRRHSSSTRSGGDLTLGLEPSEEAPRS 1073
DB 1148 PSTIAETDEYLQPKTRPSIMLPGPSA-----VEPS-DEMPKS 1183
QY 1074 -----PLAP-----SEGASDVFDDGLGMGAAGKLSLQTHDPSPLQRYSEDPVPLPSET 1124
DB 1184 LRYCKDPLKPDDETHGKEV-----GVGGIR-----LNLPLDE 1217
QY 1125 DGYVAPLTCSPQPEYVNOFVRPQPPSPREGPLPAARPAAGATLERAKTLSFGKNGVWQDV 1184
DB 1218 DDYLMF-TCSQ-----NQS-----TFG-----YMDL 1238
QY 1185 FAFGAVENPEYL-----TPGGAAPOHPPPAPFSPAPDNLYYWDQPPPERGAPPS 1235
DB 1239 IGVPAVDNPEYLMGSTQAIAGLAQSGMG---PHTPP-----PPN 1275
QY 1236 TFKGTPTAENPE 1247
DB 1276 TPNGMPTHQHSQ 1287
```

RESULT 8

Q9UK79

ID Q9UK79 PRELIMINARY; PRT; 419 AA.

AC Q9UK79;

```
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBAJ databases.
DR EMBL: AF177761; AAD56009.2; -
DR InterPro: IPR000494; EGFR_L domain.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR SMART: SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 7.2e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPFGAASQVCTGTDMKRLRPASPTHLDMLRHLQGCQVQGNL 60
DB 1 MELAALCRWGLLLALLPFGAASQVCTGTDMKRLRPASPTHLDMLRHLQGCQVQGNL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLQEDNVALAVLDNG 120
DB 61 ELTYLPTNASLFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLQEDNVALAVLDNG 120
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNQLA 180
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCOSLTRVCAGGCARCGPLPTDCCHQEC 240
DB 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCOSLTRVCAGGCARCGPLPTDCCHQEC 240
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGL 345
DB 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVTHSL 345

RESULT 9
Q9R2X1 PRELIMINARY; PRT; 367 AA.
ID Q9R2X1
AC Q9R2X1;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBAJ databases.
```

```
DR EMBL; BC027080; AA27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.5e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESILRRRFTTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICT 948
Db 1 MALESILRRRFTTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICT 60

Qy 949 IDVTMIMVKWIMIDSECRPRELVSFSESMARDPQRVFIQNEIDLGPASLDSTFYRSL 1008
Db 61 IDVTMIMVKWIMIDSECRPRELVSFSESMARDPQRVFIQNEIDLGPSSPMDSTFYRSL 120

Qy 1009 LEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTSRSGGDLTLGLEPSEE 1068
Db 121 LEDDMDGLVDAEYLVPOQGFPCDPALGTGSTAHRHRSSTSRSGGDLTLGLEPSEE 180

Qy 1069 EAPRSPLAPSEGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDTVPLPSETDGYV 1128
Db 181 EAPRSPLAPSEGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDTVPLPSETDGYV 240

Qy 1129 APLTCSPOPEVNOPDVRRPPSPREGPLPAAPAGATLERAKTSLSPKNGVWKFVAFG 1188
Db 241 APLTCSPOPEVNOPDVRRPPSPREGPLPAAPAGATLERAKTSLSPKNGVWKFVAFG 300

Qy 1189 GAVENPEYLVPOQGAAPQHPPPAFSPAFONLYYWDQDPPRGAAPPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLVPOQGAAPQHPPPAFSPAFONLYYWDQDPPRGAAPPSTFKGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;
```

```
Query Match      25.3%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.2e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;

Qy 569 PQNGSVTCFGEADOCVACAHYKDPFPCVACRCPGVKPDLSYMPILWKEPDEEGACQCP 628
Db 141 PEETATPKTGP--DHCWKCAHFDGPHCVKACPAVIGENDTL-VWKYADANAVCOLCHP 197

Qy 629 NCHSCVDLDDKGPACQASPLTSIVSAVV-GILLVVVLGVFGILIKRQOKIRKYTM 687
Db 198 NCTRGCKGPGLEGCP--NGSKTPSIAAGVVVGLLCLVVVGLGIGLYLRRR-HIVRKR 253

Qy 688 RLLQETELVEPLTPSCAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIDGENVK 747
Db 254 RLLQERELVEPLTPSGEAPNQAHRLIKETEFKVKVVLGSGAGFTVYKGLWIPEGEVK 313

Qy 748 IPVAIKVLRNTSPKANKEILDEAYVNAVGVSPVSRLLGICLTSTVOLVTLQMPYGC 807
Db 314 IPVAIKELRATSSPANKEILDEAYVNAVSDNPRVCRLLGICLTSTVOLVTLQMPYGC 373

Qy 808 DHVRENRLGSDLLNWCMIQIAKGMSTYLEDVRLVHRDLAARNVLKSPNHVKTDFGLA 867
Db 374 DYIREHKDNIGSQVLLNWCMIQIAKGMNLEERRLVHRDLAARNVLVKTPOHVKTDFGLA 433

Qy 868 RLLDIDETEHADGKVPKIMMALESILRRRFTTHOSDWSYGVTVWELMTGAKPYDGIP 927
Db 434 KLLGADEKEYHAEGKVPKIMMALESILHRIYTHOSDWSYGVTVWELMTGSKPYDGIP 493

Qy 928 AREIPDLEKGERLPOPICTIDVYMTIMVKWIMIDSECRPRELVSFSESMARDPQRV 987
Db 494 ASELSVLEKGERLPOPICTIDVYMTIMVKWIMIDSECRPRELVSFSESMARDPQRV 553

Qy 988 VIQ-NEDLGPASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFPCDPAPGAGMVHHR 1046
Db 554 VIQDERMHLPSPTDSKFYRTLMEEDMEDIVDAEYLVPHQGF----- 598

Qy 1047 HRSSTSRSGGDLTLGLEPSEEEAPRSP-----APSEGAGSVDFDGLGMAAKGLOS 1101
Db 599 -NSPST-----SRTPLSSLSATSNNSATNCID-----RNGQGH 631

Qy 1102 PTHDPSPLQRYSEDTVPLPSET--DCGVAPLTCSPOPEVNOPDVRRPPSPREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKFPS----- 675

Qy 1160 ARPAGATLERAKTSLSPKNGVWKFV-----AFGGAVENPEYL 1197
Db 676 -----TAMVQNIYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
```





```

Qy 996 PASPLDSTFYRSLLEDDMDGLVDAEYLVPOQGFCCPDAPGAGGMVHHRSSSTRSG 1055
Db 414 LPSPTDSKFKYRTLMEEDMEDIVADSYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTLCLESEAEAPRSL-----APSEGAGSDVFGDLMGNAKGLQSLPTHDPSPLO 1110
Db 455 -----SKTPLLSSLSATSNNSNCIDRNG-----H----- 481
Qy 1111 RYSEDPVTPLPSETDGVVAPLTCSPQPEYVQNPQVRPQPPSPREGPLPAARAGAT-LE 1169
Db 482 -----PVREDGFL-----PAPEYVQ--LMPKKSTAMVQIQIYNYISLTAKS 523
Qy 1170 AKTLPSPKGVVQKDVFAFGAVENPEYL 1197
Db 524 LPMDSRYN-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
(1)
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maithe N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
(2)
RN SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7AC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maithe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
(3)
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuoka H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Sakurai L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bareh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayaishizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.7%; Score 1476.5; DB 11; Length 655;
Best Local Similarity 43.9%; Pred. No. 8.8e-103;
Matches 281; Conservative 97; Mismatches 237; Indels 25; Gaps 9;

Qy 11 LLLALLPPGAA--STOVCTGTDMKRLPASPEHDLMLRHLVQGCQVVOGNLELYLPTN 68
Db 14 LTTALCAAGALBEKKVCGQTSNRLTQLGTFEDHFLSLQRMNVCVWLGNLLEITYVQRN 73

Qy 69 ASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRVGTQLFEDNVALAVLDNGDPLNNTTP 128
Db 74 YDLSFLTKIOEVAGYVLIHNTVERIPLENLQIRGNALYENTYALAILSN----- 124

Qy 129 VTGASPGGLRELQRLSILTEILKGVLIQRNPOLCYDTILWKDI----FHKNNQLALTLI 184
Db 125 -YGNRTGLRELPMRLNQLLEILGAVFSNNPILCNMDTIQWRDIQNVVFSNMSMDL--- 180

Qy 185 DTRNSRACHPCSPMKSGRCWGESSEDCOSLRTVTCAGGCA-RCKGPLPTDCCHCEQCAAG 243
Db 181 -QSHPSSCPCKDPCPNCSGWGGEENCOKLTKIIQAQCSHRCGRSPSDCCCHNQCAAG 239

Qy 244 CTGPKHSDCLCLHFNHSGICELHCPALTYNTDTFESHNPGRVTFGASCVTACPNY 303
Db 240 CTGPRESDECLVCOKFODEATCKDTCPLMLYNFTTYQMDVNPBGKVSFGATCKKCPRY 299

Qy 304 LSTDVGSCTLVCPHNRQEVTAEDGTQRCBCKSKPCARVCYGLQMVYKANSKFGITELE 363
Db 300 VTDHGSCVACGPDYEV-EEDGIRKCKKCDGCPKVCNGIGIGBEFK-DTSLINATNIK 357

Qy 364 -FAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEITGYLISAMPDSDL 422
Db 358 HFKYCTAISGDLHLPLVAFKGSFTRTPPLDPRELILKTVKEITGFLLIQAMPDNTDL 417

Qy 423 SVFQNLQVIRGRILHNGATSLTQGLGISWGLRSRLBELGSLALHHNTHLCFVHTVPW 482
Db 418 HAFENLEIRGRTKQHQGFSLAVVGLNITSLGRSLKEISDGDVVIISGNRLCYANTINW 477

Qy 483 DQLFROYIKANSKFIGITELEVCVEGLACHQLCARGHCWGPPTQCVNCSQFRLRGSCVE 542
Db 478 KKLFGTPTNOKTKIMNRAEKDKAVNVHVCNPLCSSGCGWGPEDPCVCSQNVSRGECVE 537

Qy 543 ECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFQPEADQCVCACAHYKDPFPCVARCP 602
Db 538 KCVILEGEPREFVENSECICQHEPCLQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPA 597

Qy 603 GVKPDLSTYMPIWKFPDEEAGCQPCPNCTHSCVDLDDKGC 642
Db 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGC 636

```

Search completed: July 22, 2003, 09:01:22  
Job time : 53.3575 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds  
(without alignments)  
4403.399 Million cell updates/sec

Title: SEQ4-465-479-12  
Perfect score: 6804  
Sequence: 1 MELAALCRWGLLLALLPPCA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6637	97.5	1255	21	AA1992620 Human heregulin 2
2	6637	97.5	1255	22	AAE12130 Human tyrosine kin
3	6637	97.5	1255	22	AAE60167 HER2 transgene pla
4	6637	97.5	1255	23	AAU74545 Human HER2 (ErBB2)
5	6631	97.5	1255	17	AAW01111 HER-2/neu protein
6	6631	97.5	1255	20	AAW92406 Human HER-2/neu pr
7	6631	97.5	1255	21	AAW92406 Human HER-2/neu pr
8	6631	97.5	1255	21	AAW84780 Amino acid sequenc
9	6631	97.5	1255	22	AAW84548 Human HER-2/neu pr
10	6631	97.5	1255	22	AAW88267 HER2/neu amino aci

11	6631	97.5	1255	23	AAE24067 Human Her-2 protei
12	6631	97.5	1255	23	AAE20479 Human Her-2/neu pr
13	6631	97.5	1255	23	AAW51143 Human Her-2/neu on
14	6631	97.5	1255	23	AAU77114 Human Her-2/neu po
15	6588	96.8	1433	14	AAK39568 Sequence of c-erbB
16	6467	95.0	1223	23	AAU98923 Human breast cance
17	6314	92.8	1200	21	AAE21208 Human HER-2/neu pr
18	5854.5	86.0	1256	21	AAE21199 Rat HER-2/neu prot
19	5854.5	86.0	1256	23	AAW51144 Rat HER-2/neu onco
20	5827.5	85.6	1256	21	AAE21206 Mouse Her-2/neu pr
21	5827.5	85.6	1256	22	AAE21206 Amino acid sequenc
22	5827.5	85.6	1256	23	AAW51151 Mouse Her-2/neu on
23	4717	69.3	919	21	AAE21203 Human HER-2/neu fu
24	4717	69.3	919	23	AAW51148 Her-2/neu extracel
25	3977.5	58.5	920	23	AAW51152 Mouse Her-2/neu ex
26	3977.5	58.5	926	23	AAW51153 Mouse Her-2/neu ex
27	3601	52.9	712	21	AAE21204 Human HER-2/neu fu
28	3601	52.9	712	23	AAW51149 Her-2/neu extracel
29	3455	50.8	782	18	AAW19764 Her2-GM-CSF immuno
30	3453	50.7	653	21	AAE21200 Extracellular HER-
31	3453	50.7	653	23	AAW51145 Human Her-2/neu on
32	3415	50.2	645	22	AAE60408 Human ErbB2 oncopr
33	3415	50.2	645	22	AAE61593 Human ErbB2 extrac
34	3350	49.2	951	21	AAU44993 DC86FV-erbB2EC fu
35	3247	47.7	624	11	AAE08222 Extracellular port
36	3121	45.9	1210	21	AAE19259 Amino acid sequenc
37	3121	45.9	1210	21	AAE50616 Human EGF receptor
38	3121	45.9	1210	23	AAE23019 Human Her-1 protei
39	3121	45.9	1210	23	AAW50768 Human epidermal gr
40	3119	45.8	1210	22	AAE68420 Amino acid sequenc
41	3084	45.3	583	23	AAE20483 Human protein for
42	3084	45.3	587	23	AAE20481 Human protein for
43	3083	45.3	589	23	AAE20484 Human protein for
44	3083	45.3	600	23	AAE20482 Human protein for
45	3080	45.3	1210	23	ABP51768 Human epidermal gr

#### ALIGNMENTS

RESULT 1  
A\_Y92620  
ID AA1992620 standard; Protein; 1255 AA.  
XX  
AC AA1992620;  
XX  
DT 10-AUG-2000 (first entry)  
XX  
DE Human heregulin 2 (Her2).  
XX  
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
KW self-protein; cancer; breast cancer; prostate cancer;  
KW cell-associated peptide antigen; foreign epitope.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Domain  
FT Location/Qualifiers  
FT 1..173  
FT /label= N-terminal  
FT /note= "mature polypeptide"  
FT 5..25  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 59..73  
FT /label= insertion\_region  
FT /note= "suitable for foreign epitope insertion"  
FT 103..117  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 149..163  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 174..323  
FT Domain





QY 781 YVSRLLGICLTSTVOLTPMLPYGCLLDHVRNRRGLSGODLLNWKMQIAKMSYLEVDYR 840  
Db |||||  
QY 781 YVSRLLGICLTSTVOLTPMLPYGCLLDHVRNRRGLSGODLLNWKMQIAKMSYLEVDYR 840  
Db |||||  
QY 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900  
Db |||||  
QY 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900  
Db |||||  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPQPPICITIDVYMIWVKWM 960  
Db |||||  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLLEKGERLPQPPICITIDVYMIWVKWM 960  
Db |||||  
QY 961 IDSECRPRFRELVSFSEMRADPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMDGLVDA 1020  
Db |||||  
QY 961 IDSECRPRFRELVSFSEMRADPQRFVVIQNEIDLGASPLDSTFYRSILLEDDMDGLVDA 1020  
Db |||||  
QY 1021 EYLVPOQGFCCPDPAAGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db |||||  
QY 1021 EYLVPOQGFCCPDPAAGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db |||||  
QY 1081 AGSDVFDGDLGMAAGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db |||||  
QY 1081 AGSDVFDGDLGMAAGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db |||||  
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPKNGVVKDVPAFCGAVENPEYLTPO 1200  
Db |||||  
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPKNGVVKDVPAFCGAVENPEYLTPO 1200  
Db |||||  
QY 1201 GGAAPQHPHPPAFSPAFDNLYWDDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db |||||  
QY 1201 GGAAPQHPHPPAFSPAFDNLYWDDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db |||||

RESULT 2  
AAE12130  
ID AAE12130 standard; Protein; 1255 AA.  
XX  
AC AAE12130;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human tyrosine kinase-type receptor, HER-2.  
XX  
KW Therapeutic compound; major histocompatibility complex; vaccine;  
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;  
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
KW antigen presenting cell; human; tyrosine kinase-type receptor.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 774..782  
FT /note= "Antigenic epitope"

XX  
XX WO200168677-A2.  
XX  
XX 20-SEP-2001.  
XX  
XX 16-MAR-2001; 2001WO-US40328.  
XX  
XX 16-MAR-2000; 2000US-0527487.  
XX  
XX (GENZ ) GENZYME CORP.  
XX  
XX Nicolette CA;  
XX  
XX WPI; 2001-616284/71.  
XX  
XX N-PSDB; AAD19731.  
XX

PT Novel synthetic therapeutic compound for inducing immune response and  
PT for use in adoptive immunotherapy, has enhanced binding to major  
PT histocompatibility molecules and enhanced immunoregulatory properties  
PT

XX  
PS Claim 4; Page 63-67; 69pp; English.  
XX  
CC The invention relates to synthetic therapeutic compounds (antigenic  
CC peptides) with enhanced binding to major histocompatibility complex  
CC (MHC) molecules and enhanced immunoregulatory properties relative  
CC to their natural counterparts. Compounds of the invention are useful  
CC for inducing an immune response in a subject and for use in adoptive  
CC immunotherapy. They are useful as components of anti-cancer vaccines  
CC and to expand immune effector cells that are specific for cancers  
CC characterised by expression of the breast cancer antigen, HER-2.  
CC Polynucleotides that encode peptides of the invention are useful as  
CC hybridisation probes and as primers for the detection of genes of gene  
CC transcripts that are expressed in antigen presenting cells (APCs), to  
CC confirm transduction of polynucleotides into host cells. The present  
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
CC of the invention are designed based on the HER-2 antigenic peptide  
CC (774-782).  
XX  
SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6637; DB 22; Length 1255;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 1227; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLRLPASPTHLMRLHYQSCVQVQGNL 60  
Db |||||  
QY 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLRLPASPTHLMRLHYQSCVQVQGNL 60  
Db |||||  
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNQVRQVPLQRLIRVGTQLFEDNVALVDNG 120  
Db |||||  
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNQVRQVPLQRLIRVGTQLFEDNVALVDNG 120  
Db |||||  
QY 121 DPLNNTTPVTGASPGGLRELQRLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180  
Db |||||  
QY 121 DPLNNTTPVTGASPGGLRELQRLRSITELKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180  
Db |||||  
QY 181 LTLIDTNRSRACHPCSPMKGSRGWGSESSQCSLRTVVCAGGCARCKGPLPTDCHEQC 240  
Db |||||  
QY 181 LTLIDTNRSRACHPCSPMKGSRGWGSESSQCSLRTVVCAGGCARCKGPLPTDCHEQC 240  
Db |||||  
QY 241 AAGCTGPKHSCLACLHPNHSIGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300  
Db |||||  
QY 241 AAGCTGPKHSCLACLHPNHSIGICELHCPALVYNTDTFESMPNPEGRTYFGASCVTACP 300  
Db |||||  
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db |||||  
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360  
Db |||||  
QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFFETLEITGYLISANPDSL 420  
Db |||||  
QY 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFFETLEITGYLISANPDSL 420  
Db |||||  
QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSRLRELSGLALIHNNTHLCFVHTV 480  
Db |||||  
QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSRLRELSGLALIHNNTHLCFVHTV 480  
Db |||||  
QY 481 PWDOLFRQYIKANSKFIGITELCEVCGSLACHQICARCHGKPGPTQCVCNSQFLRGQEC 540  
Db |||||  
QY 481 PWDOLFRNPHQALLHTANRPEDECVGEGSLACHQICARCHGKPGPTQCVCNSQFLRGQEC 540  
Db |||||  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEVCAR 600  
Db |||||  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEVCAR 600  
Db |||||  
QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSVYG 660  
Db |||||  
QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSVYG 660  
Db |||||  
QY 661 ILLVVLGVVFGILIKERQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Db |||||  
QY 661 ILLVVLGVVFGILIKERQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Db |||||

Qy 721 RKVKLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Dd  
Qy 721 RKVKLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Dd  
Qy 781 YVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRRGLSGDQLLNKMCQIAKMSYLEVDV 840  
Dd 781 YVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRRGLSGDQLLNKMCQIAKMSYLEVDV 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
Dd 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960  
Dd 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVKCM 960  
Qy 961 IDSECRPRFRELSEFSSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDGDLVDA 1020  
Dd 961 IDSECRPRFRELSEFSSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDGDLVDA 1020  
Qy 1021 EYLVPOQGFPCPPAPGAGMWHRRSSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080  
Dd 1021 EYLVPOQGFPCPPAPGAGMWHRRSSSTRSGGDLTLGLSPSEEAAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
Dd 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSGKNGVVDVPAFGAVENPEYLTPO 1200  
Dd 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSGKNGVVDVPAFGAVENPEYLTPO 1200  
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVPU 1255  
Dd 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVPU 1255

RESULT 3  
AAB60167  
ID AAB60167 standard; Protein; 1255 AA.  
AC AAB60167;  
XX  
DT 03-APR-2001 (first entry)  
XX  
DE HER2 transgene plasmid construct encoded protein.  
XX  
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
KW antibody.  
XX  
OS Homo sapiens.  
OS Synthetic.  
XX  
PN WO200100244-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-US17229.  
XX  
PR 25-JUN-1999; 99US-0141316.  
PR 16-MAR-2000; 2000US-0189844.  
XX  
FA (GETH ) GENENTECH INC.  
XX  
PI Erickson S, Schwall R;  
XX  
DR WPI; 2001-061962/07.  
DR N-PSDB; AAF24297.  
XX  
PT Treating tumors, particularly breast cancers, which overexpress an ErbB  
PT receptor and does not respond to an anti-ErbB antibody, comprises  
PT conjugating the antibody to a maytansinoid -  
XX

PS Example 3; Fig 4; 92pp; English.  
XX  
CC The present invention provides a method of treating cancer by  
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In  
CC particular, the antibody is directed against ErbB2 (also known as HER2  
CC and p185neu). The method is particularly useful in the treatment of  
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,  
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.  
XX  
SQ Sequence 1255 AA;  
  
Query Match 97.5%; Score 6637; DB 22; Length 1255;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 1227; Conservative 5; Mismatches 23; Indels 0; Gaps 0;  
  
Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGCCVVOGNL 60  
Dd 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGCCVVOGNL 60  
Qy 61 ELTYLPTNASLFLQDIEVOGYVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120  
Dd 61 ELTYLPTNASLFLQDIEVOGYVLIAHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120  
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQLA 180  
Dd 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQLA 180  
Qy 181 LTLDTNRSRACHPCSMCKGSRGWSSESDECSQSLTRTVGAGGCARCKPLTDCCHQC 240  
Dd 181 LTLDTNRSRACHPCSMCKGSRGWSSESDECSQSLTRTVGAGGCARCKPLTDCCHQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Dd 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPLHNQVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360  
Dd 301 YNYLSTDVGSCTLVCPLHNQVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360  
Qy 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLQPELOQVFTLEITGYLISAWPDSL 420  
Dd 361 IQEFAGCKKIFGSLAFIPESPDGPASNTAPLQPELOQVFTLEITGYLISAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGSLWGLRSLRSLGSLALTHHNLFCVHTV 480  
Dd 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGSLWGLRSLRSLGSLALTHHNLFCVHTV 480  
Qy 481 PWDQLFROYIKANSKFITGITELECVGEGLAGHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Dd 481 PWDQLFRNPHQALLTANRPEDECVGEGLAGHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600  
Dd 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600  
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTISIVSAVVG 660  
Dd 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTISIVSAVVG 660  
Qy 661 ILLVVVLGVVFGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Dd 661 ILLVVVLGVVFGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Dd 721 RKVKLGSGAGTGYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Qy 781 YVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRRGLSGDQLLNKMCQIAKMSYLEVDV 840  
Dd 781 YVSRLLGICLTSTVOLVTQLMPYGLLDHVRNRRGLSGDQLLNKMCQIAKMSYLEVDV 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
Dd 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900

Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
 Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWCWM 960  
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVWCWM 960  
 Qy 961 IDSECRPRFRELVEFSRMDRPFVVVIONEDLGPASPLDSTFYRSLLEDGDLVDA 1020  
 Db 961 IDSECRPRFRELVEFSRMDRPFVVVIONEDLGPASPLDSTFYRSLLEDGDLVDA 1020  
 Qy 1021 EYLVPQOQFFCFDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
 Db 1021 EYLVPQOQFFCFDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
 Qy 1081 AGSDVFDDGLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
 Db 1081 AGSDVFDDGLGMAAKGLQSLPHDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
 Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKGVVKKVFAFGGAVENPEYLTPO 1200  
 Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPGKGVVKKVFAFGGAVENPEYLTPO 1200  
 Qy 1201 GGAAPQHPPPAFSPAFDNLVYDQDPPPERGAPPSTFKGTPAENPEYLGLDVVPV 1255  
 Db 1201 GGAAPQHPPPAFSPAFDNLVYDQDPPPERGAPPSTFKGTPAENPEYLGLDVVPV 1255

## RESULT 4

AAU74545

ID AAU74545 standard; Protein; 1255 AA.

AC AAU74545;

XX

DT 23-APR-2002 (first entry)

DE Human HER2 (ErbB2) polypeptide.

XX Human: HER2; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glial disorder; astrocytal disorder; hypothalamic disorder;  
 KW glandular disorder; macrophagal disorder; epithelial disorder;  
 KW atromal disorder; blastocoealic disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

OS US2002001587-A1.

PN 03-JAN-2002.

PD 16-MAR-2001; 2001US-0811123.

PF 16-MAR-2001; 2000US-189844P.

PR 05-OCT-2000; 2000US-238327P.

PX (ERIC/) ERICKSON S.

PA (SCHW/) SCHWALL R.

PA (SLIW/) SLIWKOWSKI M.

XX Erickson S, Schwall R, Sliwkowski M;

PI WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

PT Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.

XX SQ Sequence 1255 AA;

Query Match 97.5%; Score 6637; DB 23; Length 1255;

Best Local Similarity 97.8%; Pred. No. 0;

Matches 122; Conservative 5; Mismatches 23; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMKRLPASPEHLDMLRHLVQCCVQVQNL 60  
 Db 1 MELAALCRWGLLLALLPFGAASSTQVCTGTDMKRLPASPEHLDMLRHLVQCCVQVQNL 60  
 Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVRQVPLQRLRIVRGCTQLFEDNYALAVLDNG 120  
 Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNNQVRQVPLQRLRIVRGCTQLFEDNYALAVLDNG 120  
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180  
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNQOLA 180  
 Qy 181 LTLIDNRSRACHPCSPMKCGSRGWESSEDQSLTRTVCAAGCARCKGPLTDCCHQEC 240  
 Db 181 LTLIDNRSRACHPCSPMKCGSRGWESSEDQSLTRTVCAAGCARCKGPLTDCCHQEC 240  
 Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Qy 301 YNYLSTDVSGCTLVCPHNNQVTAEDGTORCEKSKPCARVCYGLGHQYIKANSKFGIT 360  
 Db 301 YNYLSTDVSGCTLVCPHNNQVTAEDGTORCEKSKPCARVCYGLGHQYIKANSKFGIT 360  
 Qy 361 ELEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEEITGYLYISAWPSLP 420  
 Db 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEEITGYLYISAWPSLP 420  
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSLRELGSGLALIHNNTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSLRELGSGLALIHNNTHLCFVHTV 480  
 Qy 481 PWDOLFROYIKANSKFITIGITELCEVGEGLACHQICARGHCWGPGPTQCVNCSQFLRQEC 540  
 Db 481 PWDOLFQFNHQAALLHTANRPEDECVGEGLACHQICARGHCWGPGPTQCVNCSQFLRQEC 540  
 Qy 541 VEBCRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFPCVARC 600  
 Db 541 VEBCRVLQGLPREYVVARHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFPCVARC 600  
 Qy 601 PGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660  
 Db 601 PGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTSIVSAVVG 660  
 Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780  
 Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETLDEAYVMAGVGP 780  
 Qy 781 YVSRLLGICLTSTVQLVTQTMPYGCILLDHVRENRLGSGQDLNLNWCQIAKSGMYSLEDVR 840

Db 781 YVSRLLGICLTSTVQLVTPMPYGLCLDVRNRLGSLQDLNWMQIAKMSYLEDR 840  
QY 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900  
Db 841 LVHRDLARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRPT 900  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITDVYIMVWKWM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITDVYIMVWKWM 960  
QY 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLVDA 1020  
Db 961 IDSECRPRFRELVSFSESMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLVDA 1020  
QY 1021 EYILVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
Db 1021 EYILVPOQGFCCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
QY 1081 AGSDVDFDGLGMAAGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVDFDGLGMAAGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKGVVVDVPAFGAVENPEYLTPO 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKGVVVDVPAFGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5  
AAW01111  
ID AAW01111 standard; Protein; 1255 AA.  
XX AC AAW01111;  
XX 01-JAN-1997 (first entry)  
XX DE HER-2/neu protein.  
XX KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; immunisation; tumour; vaccine; vector.  
XX OS Homo sapiens.  
XX FH Key Location/Qualifiers  
FT Domain 676..1255  
FT /label= Intracellular domain  
FT /note= "claimed domain, useful for immunisation"  
XX PN W09630514-A1.  
XX XX  
PD 03-OCT-1996.  
XX XX  
PF 28-MAR-1996; 96WO-US01689.  
XX XX  
PR 31-MAR-1995; 95US-0414417.  
XX XX  
PA (UNIW ) UNIV WASHINGTON.  
XX Cheever MA, Disis ML;  
XX WPI; 1996-455361/45.  
XX N-PSDB; AAT40739.  
XX  
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or  
PT treatment of malignancies with which the HER-2/neu oncogene is  
PT associated  
XX  
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.  
XX SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6631; DB 17; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLKRLPASPTHLDMLRLHYQCQVVOGNL 60  
Db 1 MELAALCRWGLLLALLPFGAASTQVCTGTDMLKRLPASPTHLDMLRLHYQCQVVOGNL 60  
QY 61 ELTYLPTNASLSFLQDIEQVQGYVLI AHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120  
Db 61 ELTYLPTNASLSFLQDIEQVQGYVLI AHNQVRQVPLQRLRIVRGTLFEDNYALAVLNG 120  
QY 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNOLA 180  
Db 121 DPLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSILTRVCAGGCARCKGPIPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSILTRVCAGGCARCKGPIPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLPNHSIGICELHCPALVTNTDTFESMPNPEGYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLPNHSIGICELHCPALVTNTDTFESMPNPEGYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIT 360  
Db 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIT 360  
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQVFTLEBEITGYLISAWPDSLUP 420  
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQVFTLEBEITGYLISAWPDSLUP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRLSRLGSLALIHNTLHLCFVHTV 480  
QY 481 PWDOLFROYIKANSKFITILECECVGELACHQICARGHCWGPGTQCNCVSQFLRGQEC 540  
Db 481 PWDOLFROYIKANSKFITILECECVGELACHQICARGHCWGPGTQCNCVSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFVCVARC 600  
Db 541 VEECRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFVCVARC 600  
QY 601 PSYKVPDLNLYMPIWKFFDEGACQPCINCHTSCVDLDDKCPAEQASPLTSIVSAVVG 660  
Db 601 PSYKVPDLNLYMPIWKFFDEGACQPCINCHTSCVDLDDKCPAEQASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
QY 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
QY 781 YVSRLLGICLTSTVQLVTPMPYGLCLDVRNRLGSLQDLNWMQIAKMSYLEDR 840  
Db 781 YVSRLLGICLTSTVQLVTPMPYGLCLDVRNRLGSLQDLNWMQIAKMSYLEDR 840

QY 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900  
DB 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960  
QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020  
DB 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDDMGLVDA 1020  
QY 1021 EYLVVQOQFFCDPAPGACGMVHHRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080  
DB 1021 EYLVVQOQFFCDPAPGACGMVHHRSSSTRSGGDLTLGLEPSEEAAPRPLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200  
DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVYKDVFAFGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255  
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVPV 1255

RESULT 6  
AAW92406  
ID AAW92406 standard; Protein; 1255 AA.  
XX  
AC AAW92406;  
XX  
DT 21-APR-1999 (first entry)  
XX  
DE Human HER-2/neu oncogene protein.  
XX  
KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
KW malignancy; treatment; tumour.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Region 676..1255  
XX /note= "region which elicits immune response"  
XX  
PN US5869445-A.  
XX  
XX 09-FEB-1999.  
XX  
XX 01-APR-1996; 96US-0625101.  
XX  
XX 01-APR-1996; 96US-0625101.  
XX 17-MAR-1993; 93US-0033644.  
XX 12-AUG-1993; 93US-0106112.  
XX 31-MAR-1995; 95US-0414417.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
XX  
XX Cheever MA, Disis ML;  
XX WPI; 1999-152835/13.  
XX N-PSDB; AAX01912.  
XX  
XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
PT an HER-2/neu associated malignancy, particularly for treating or  
PT preventing tumours  
XX  
XX Claim 3; Column 31-38; 26pp; English.  
XX  
XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune  
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
CC B cells to produce an immune response to the HER-2/neu protein. The  
CC method can be used for immunisation against a malignancy in which the  
CC HER-2/neu oncogene is associated and in the treatment of an existing  
CC tumour, or to prevent tumour occurrence or reoccurrence.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6631; DB 20; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLQGCQVQGNL 60  
DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPTHLDMLRHLQGCQVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLIHAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLIHAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKNOLA 180  
DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIORNPOLCYODTILWKDIFHKNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCEKSPCARVCYGLGHQYIKANSKFIT 360  
DB 301 YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCEKSPCARVCYGLGHQYIKANSKFIT 360  
QY 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVPEETLEETIGLYLISAWPSLP 420  
DB 361 IQSFAGCKIFGSLAFIPESFDGDPASNTAPLOPEQLOVPEETLEETIGLYLISAWPSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRLSGSLALIHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGSLRLSGSLALIHNNTHLCFVHTV 480  
QY 481 PWDQLFRQYIKANSKFITIGITELECVGEGLACHQLCARGHCWGPGTQCVNCSQFLRGEC 540  
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVNCSQFLRGEC 540  
QY 541 VEBCRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARG 600  
DB 541 VEBCRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARG 600  
QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQSRASPLTSIVAVVG 660  
DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQSRASPLTSIVAVVG 660  
QY 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
DB 661 ILLVVLGVVFGILIKRROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEITLDEAYVMAGVSP 780  
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEITLDEAYVMAGVSP 780  
QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRLGSLQDILLNWCQIAKGSYLEDVR 840  
DB 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRLGSLQDILLNWCQIAKGSYLEDVR 840  
QY 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900  
DB 841 LVHRDLAARNVLKSPNHNKIDTDFGLARLLDIDETEHADGGKVPKWMALLESILRRRFT 900

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVCKWM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVCKWM 960  
Qy 961 IDSECRPRFRELVSFSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020  
Qy 1021 EYLVPOQGFCDPAPACAGMWHHRSSSTRSGGDLTLGLLEPSEEEAPRSPAPSEG 1080  
Db 1021 EYLVPOQGFCDPAPACAGMWHHRSSSTRSGGDLTLGLLEPSEEEAPRSPAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMAAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NQDVRPOPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200  
Db 1141 NQDVRPOPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVFAGGAVENPEYLTPO 1200  
Qy 1201 GGAAPQHPPPAFSPADNLYYMDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255  
Db 1201 GGAAPQHPPPAFSPADNLYYMDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7  
AAB21198  
ID AAB21198 standard; protein; 1255 AA.  
XX  
AC AAB21198;  
XX  
DT 12-JAN-2001 (first entry)  
XX Human HER-2/neu protein.  
XX  
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX  
OS Homo sapiens.  
XX  
PN WQ200044899-A1.  
XX  
PD 03-AUG-2000.  
XX  
PF 28-JAN-2000; 2000MO-US02164.  
XX  
PR 29-JAN-1999; 99US-0117976.  
XX  
PA (CORI-) CORIXA CORP.  
PA (SMIK ) SMITHKLINE BEECHAM.  
XX  
PI Cheever MA, Gheysen D;  
XX  
DR WPI; 2000-505976/45.  
XX  
PS N-PSDB; AAA89736.  
XX  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX  
PS Claim 52; Fig 7; 128pp; English.  
XX  
CC The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6631; DB 21; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTOMKRLPASPTHLDMLRHLVGGCVVQGNL 60  
Db 1 MELAALCRWGLLALLPPGAASQVCTGTOMKRLPASPTHLDMLRHLVGGCVVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVROVPLQRLRIVRGTLQFEDNYALAVLNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHAHQVROVPLQRLRIVRGTLQFEDNYALAVLNG 120  
Qy 121 DPLNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHNNOLA 180  
Db 121 DPLNTTPTVGTASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDIFHNNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCSLTRTVCAAGCARCKGPLTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCSLTRTVCAAGCARCKGPLTDCCHQC 240  
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPLHNOEVTAEQTCRCKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLHNOEVTAEQTCRCKSKPCARVCYGLGMEHLREVRAVTSAN 360  
Qy 361 ELEFACKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGYLYISAWPDSL 420  
Db 361 IQFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIGYLYISAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRIILHNGAYSLTLQGLGISWLGRLSLRELGLSLALIHNTLHLCFVHT 480  
Db 421 DLSVFQNLQVIRGRIILHNGAYSLTLQGLGISWLGRLSLRELGLSLALIHNTLHLCFVHT 480  
Qy 481 PWDQLFRQYIKANSKFIGITELSCVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGOEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGOEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVCAHYKDPFCVARC 600  
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660  
Qy 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPGAMPNQAMRILKETEL 720  
Db 661 ILLVVLGVVFGILIKRQOKIRKTYMRRLLQETELVEPLTPGAMPNQAMRILKETEL 720  
Qy 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDEAYVWAGVSP 780  
Db 721 RKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDEAYVWAGVSP 780  
Qy 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGDQLNWCQIAGMSYLEVDV 840  
Db 781 YVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGDQLNWCQIAGMSYLEVDV 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVCKWM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVCKWM 960  
Qy 961 IDSECRPRFRELVSFSESRMARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLVDA 1020

|||||  
Db 961 IDSECRPRFELVSEFSRMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
QY 1021 EYLVPQQGFCFDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db 1021 EYLVPQQGFCFDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLORYSEDPVPLPSETDGVVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLORYSEDPVPLPSETDGVVAPLTCSPQPEYV 1140  
QY 1141 NQDVPAPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200  
Db 1141 NQDVPAPQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVVDVFAFGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 8

AAV84780  
ID AAY84780 standard; Protein; 1255 AA.  
XX  
AC AAY84780;  
XX  
DT 08-AUG-2000 (first entry)  
XX  
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.  
XX  
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KW bone resorption; inflammatory disease; degenerative disorder;  
KW wound healing.  
XX  
OS Homo sapiens.  
XX  
PN WO200020579-A1.  
XX  
PD 13-APR-2000.  
XX  
PF 01-OCT-1999; 99WO-CA00912.  
XX  
PR 02-OCT-1998; 98US-0165192.  
XX  
PA (UYMC-) UNIV MCMASTER.  
PI Muller WJ, Siegel PM;  
PI WPI; 2000-303768/26.  
DR N-PSDB; AAA14812.  
XX  
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE  
PT erbB-2, inhibitors of the protein are useful for treatment of cancer -  
PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The  
CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
CC conserved cysteine residues, compared to the unspliced protein. The  
CC erbB-2 polynucleotide is used to construct probes for detecting  
CC disorders of cell transformation such as cancer. Antibodies to the  
CC protein may be used to detect SPLICE erbB-2 in a sample. Agents  
CC (e.g. antisense oligonucleotides) which inhibit the expression of  
CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and  
CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful  
CC for treating conditions involving damaged cells including conditions  
CC in which degeneration of tissue occurs, such as arthropathy, bone  
CC resorption, inflammatory diseases, degenerative disorders of the  
CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 21; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;  
QY 1 METAALCRWGLLALLPPGAASTQVCTGTOMKLRLPASPTHLDMLRHLVGGCVVQGNL 60  
Db 1 METAALCRWGLLALLPPGAASTQVCTGTOMKLRLPASPTHLDMLRHLVGGCVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIEVQGYVLIHAHQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIEVQGYVLIHAHQVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120  
QY 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180  
Db 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMCKSGSCWGESSEDCOSLTRTVCAGGCARCKGPLPTCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKSGSCWGESSEDCOSLTRTVCAGGCARCKGPLPTCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSHGI CELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLHNOEVTABDGTQRCCKSPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLHNOEVTABDGTQRCCKSPCARVCYGLGMEHLREVRVTSAN 360  
QY 361 ELRFACCKIFGSLAFELPESFDGDPASNTAPLOEQLOVFETLEEITGYLISAWPSLSP 420  
Db 361 IQEFACCKIFGSLAFELPESFDGDPASNTAPLOEQLOVFETLEEITGYLISAWPSLSP 420  
QY 421 DLSVFQNLQVIRGRIILHNGAYSUTLQGLGISWLGSLRLSLGSLALIHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRIILHNGAYSUTLQGLGISWLGSLRLSLGSLALIHNNTHLCFVHTV 480  
QY 481 PWDQLFRQYIKANSKFIGITELECVGEGEGLACHOLCARGHCWGPGPTQCVNCSQFLRGBC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGEGLACHOLCARGHCWGPGPTQCVNCSQFLRGBC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPQNGSVTCFGEADQCAHAHYKPPFCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPQNGSVTCFGEADQCAHAHYKPPFCVARC 600  
QY 601 PSGVKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSPYMPKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
QY 721 RKVKVTLGSGAFGTGYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780  
Db 721 RKVKVTLGSGAFGTGYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780  
QY 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVRENRGLGSDLLNWCQIAKGSYLEDVR 840  
Db 781 YVSRLLGICLTSTVQLVTOLMPYGLLDHVRENRGLGSDLLNWCQIAKGSYLEDVR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESLIRRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESLIRRRFT 900  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVKCWM 960  
QY 961 IDSECRPRELVSEFSRWARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
Db 961 IDSECRPRELVSEFSRWARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
QY 1021 EYLVPQQGFCFDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080



Db 1021 EYLVPQGGFFCPDPAPGAGMVRHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVYVAPLTCSPQPEYV 1140  
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
Qy 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPPERGAPSTFKGTPAENPEYLGLDVY 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPPERGAPSTFKGTPAENPEYLGLDVY 1255

## RESULT 9

AAB85458  
ID AAB85458 standard; Protein; 1255 AA.

XX AC AAB85458;

XX DT 25-SEP-2001 (first entry)

XX DE Human HER-2/neu protein.

XX KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;  
oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

XX OS Homo sapiens.

XX PN MO200153463-A2.

XX PD 26-JUL-2001.

XX PF 19-JAN-2001; 2001WO-US01850.

XX PR 21-JAN-2000; 2000US-0177545.

XX PA (CORI-) CORIXA CORP.

XX PI Cheever MA, Hand-Zimmermann S;

XX DR WPI; 2001-476112/51.

XX DR N-PSDB; AAH23392.

XX PT New antigen-presenting cells, useful as vaccines for eliciting or  
enhancing an immune response to HER-2/neu protein, particularly useful  
for treating or preventing cancer, e.g. breast cancer

XX PS Claim 2; Page 41-46; 49pp; English.

XX CC The invention provides an isolated antigen-presenting cell, which  
expresses at least an immunogenic portion of a polypeptide that produces  
an immune response to HER-2/neu protein. The antigen-presenting cells are  
useful as vaccines for eliciting or enhancing an immune response to  
HER-2/neu protein, particularly in treating or preventing malignancies in  
which the HER-2/neu oncogene is associated. Specifically, these are  
useful for treating or preventing cancer, e.g. breast cancer, ovarian,  
colon, lung or prostate cancers. The present sequence represents  
the human HER-2/neu protein (also known as p185 or c-erbB2).

XX SQ Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 22; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQTCTGTDKMLRLPASPEHLDMLRHLYQGCGVQGNL 60

Db 1 MELAALCRWGLLALLPPGAASQTCTGTDKMLRLPASPEHLDMLRHLYQGCGVQGNL 60

Qy 61 ELTYLPTNASLFLQDIEQVGYVLIHNRQVRQVPLQRLIRVRCQTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLFLQDIEQVGYVLIHNRQVRQVPLQRLIRVRCQTQLFEDNYALAVLDNG 120  
Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180  
Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDI FHKNNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKGSRGSCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGSCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPLNHNOEVAEDCTORCEKSKPCARVCYGLGWOYIKANSKFIT 360  
Db 301 YNYLSTDVGSCTLVCPLNHNOEVAEDCTORCEKSKPCARVCYGLGMEHLREAVRVTGAN 360  
Qy 361 ELSEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETTCGLYLSAWPDSL 420  
Db 361 IQEFAGCKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETTCGLYLSAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWLGSLRSLRGLGLALIHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWLGSLRSLRGLGLALIHNNTHLCFVHTV 480  
Qy 481 PWDQLPROYIKANSKFITGITELECVGGLACHOLCARGHCGWGPPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDECEVGLACHQJCARGHCGWGPPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVARC 600  
Qy 601 PSGVKPDLSTWMPNWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQERASPLTTSIVAVG 660  
Db 601 PSGVKPDLSTWMPNWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQERASPLTTSIVAVG 660  
Qy 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLQETELVEPLTPSCAMPNQAMRILKETEL 720  
Db 661 ILLVVLGVVFGILIKRRQKIRKYTMRRLQETELVEPLTPSCAMPNQAMRILKETEL 720  
Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
Qy 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGODLLNWCQIAKMSYLEVDV 840  
Db 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGLSGODLLNWCQIAKMSYLEVDV 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDGLARLADIDETEVHADGKVPKWKMALESILRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDGLARLADIDETEVHADGKVPKWKMALESILRRFT 900  
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYIMVWKCM 960  
Qy 961 IDSECRPRFRELVSFESRMARDPQRFVIVIONDLGPASPLDSTFYRSLLEDGDLVDA 1020  
Db 961 IDSECRPRFRELVSFESRMARDPQRFVIVIONDLGPASPLDSTFYRSLLEDGDLVDA 1020  
Qy 1021 EYLVPOQGGFFCPDPAPGAGMVRHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGGFFCPDPAPGAGMVRHRRSSSTRSGGDLTLGLBPSBEEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMGAAGLQSLPHTDPSPLQRYSDPTVPLPSETDGVYVAPLTCSPQPEYV 1140  
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200



Db 1141 NQDVRPQPSREGPLPAARPAATLERPKTLSPGKNGUVKDVAFEGGAVENPEYLTPQ 1200  
AAG88267  
Qy 1201 GGAAPQHPPPAFSPADNLYYWDQDPPERGAPSTFKGTTAENPEYGLDVPV 1255  
Db 1201 GGAAPQHPPPAFSPADNLYYWDQDPPERGAPSTFKGTTAENPEYGLDVPV 1255

RESULT 10

AA888267  
ID AAG88267 standard; Protein; 1255 AA.

XX AAG88267;

XX 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;  
XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.

XX Homo sapiens.

XX WO200141787-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33591.

XX 10-DEC-1999; 99US-0458299.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
XX Keogh E;  
XX WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
XX cellular immune responses for the prevention and treatment of cancer -  
XX Disclosure; Page 15; 199pp; English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).  
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
CC culture in vitro and binds to a complex of an epitope (I), bound to a  
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
CC and a second epitope and the peptide is less than 50 contiguous amino  
CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising  
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic  
CC and immunostimulant activities, and can be used in vaccines. (I), (II)  
CC and (III) are useful for inducing cellular immune responses for the  
CC prevention and treatment of cancer. (I) and (II) are useful for  
CC monitoring or evaluating an immune response to a tumour-associated  
CC antigen when incubated with a T lymphocyte sample form a patient and  
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
CC based vaccines mean that immunosuppressive epitopes that may be present  
CC in whole antigens may be avoided. Selected epitopes may be combined to  
CC enhance immunogenicity. The possible pathological side effects caused by  
CC infectious agents or whole protein antigen is eliminated. The vaccine  
CC provides the ability to direct and focus an immune response to multiple  
CC selected antigens from the same pathogen. Epitope-based anti-tumour  
CC vaccines provides the opportunity to combine epitopes derived from  
CC multiple tumour-associated molecules addressing the problem of tumour-  
CC tumour variability and reducing the likelihood of tumour escape due to  
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
CC the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 22; Length 1255;

		Best Local Similarity 97.6%; Pred. No. 0; Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;						
Qy	1	MELAAALCRWGLLALLP	GAAGAASTQVCTGT	DMKRLPASPETHLDMRLHLYGCOVVQGNL	60			
Db	1	MELAAALCRWGLLALLP	GAAGAASTQVCTGT	DMKRLPASPETHLDMRLHLYGCOVVQGNL	60			
Qy	61	ELTYLPTNASLSFLQDI	QEVQGVYVLI	IAHNVQVPLQRLRIVRGTLQFEDNYALAVLDNG	120			
Db	61	ELTYLPTNASLSFLQDI	QEVQGVYVLI	IAHNVQVPLQRLRIVRGTLQFEDNYALAVLDNG	120			
Qy	121	DPLNNTTPTVGASPGGLRELQ	LSLTLKGGVLI	QRNPOLCYODTLWKDIFHKNNOLA	180			
Db	121	DPLNNTTPTVGASPGGLRELQ	LSLTLKGGVLI	QRNPOLCYODTLWKDIFHKNNOLA	180			
Qy	181	LTLDITNRSRACHPCSP	CMCKSGRCWGESSE	DCQSLTRTV	CAGGCARCKGPLTDCCHQOC	240		
Db	181	LTLDITNRSRACHPCSP	CMCKSGRCWGESSE	DCQSLTRTV	CAGGCARCKGPLTDCCHQOC	240		
Qy	241	AAGCTGPKHSDCLAC	LHFNHSGICELH	CPALVTYNTDT	FESMPNPEGRYTFGASCVTACP	300		
Db	241	AAGCTGPKHSDCLAC	LHFNHSGICELH	CPALVTYNTDT	FESMPNPEGRYTFGASCVTACP	300		
Qy	301	YNYLSTDVGSCTLV	CPHLNQEVTAED	GTQRCCKSKPCAR	VCYGLGHQYIKANSKFTGIT	360		
Db	301	YNYLSTDVGSCTLV	CPHLNQEVTAED	GTQRCCKSKPCAR	VCYGLGHQYIKANSKFTGIT	360		
Qy	361	ELEFAGCKKIFGSLA	FLPESFDGDPAS	NATAPLOEQV	PETLEETGVLVISAWPDSL	420		
Db	361	ELEFAGCKKIFGSLA	FLPESFDGDPAS	NATAPLOEQV	PETLEETGVLVISAWPDSL	420		
Qy	421	DLSVFQNLQVIRGRI	LHNGAYS	SLTLOGLGISW	LGSLRLSELGSLALIHNTLHLCFVHTV	480		
Db	421	DLSVFQNLQVIRGRI	LHNGAYS	SLTLOGLGISW	LGSLRLSELGSLALIHNTLHLCFVHTV	480		
Qy	481	PWDQLFRQYIKANSKFTIGITE	LECEVBEGGLACHOL	CARGHCWGPGPTQCVNCSQFLRGQEC	540			
Db	481	PWDQLFRQYIKANSKFTIGITE	LECEVBEGGLACHOL	CARGHCWGPGPTQCVNCSQFLRGQEC	540			
Qy	541	VEECRVLQGLPRE	VVNARHCLPCH	PECPQONGSV	TCFPGPADQCVACAHYKDPFCVARC	600		
Db	541	VEECRVLQGLPRE	VVNARHCLPCH	PECPQONGSV	TCFPGPADQCVACAHYKDPFCVARC	600		
Qy	601	PSGVKPDLSYMP	IKFPDEEGACQPCPCIN	CTHSCVDLDDKGC	PAEQASPLTISVAVG	660		
Db	601	PSGVKPDLSYMP	IKFPDEEGACQPCPCIN	CTHSCVDLDDKGC	PAEQASPLTISVAVG	660		
Qy	661	ILLVVLGVVFGILIKRQOKIRKY	TNRLLQETEL	VEPLTPSGAMPNQAMRILKETEL	720			
Db	661	ILLVVLGVVFGILIKRQOKIRKY	TNRLLQETEL	VEPLTPSGAMPNQAMRILKETEL	720			
Qy	721	RKVKVLGSGAFGVYKGIWIPD	GENVKIPVAIKVLRENT	SPKANKETL	DEAYVMAGVGP	780		
Db	721	RKVKVLGSGAFGVYKGIWIPD	GENVKIPVAIKVLRENT	SPKANKETL	DEAYVMAGVGP	780		
Qy	781	YVSRLLGICLT	SVQLVTQ	LMPIYGC	LLDHHVNRGR	LGSDLLNWCQIAGMSYLEDVR	840	
Db	781	YVSRLLGICLT	SVQLVTQ	LMPIYGC	LLDHHVNRGR	LGSDLLNWCQIAGMSYLEDVR	840	
Qy	841	LVRDLAARNVL	VKSNNHVKIT	DFGLARLLD	IDETEHADG	GVKVPKWKMALESILRRFT	900	
Db	841	LVRDLAARNVL	VKSNNHVKIT	DFGLARLLD	IDETEHADG	GVKVPKWKMALESILRRFT	900	
Qy	901	HOSDVMWSYGVYVWELMT	FGAKPYDGI	PARBIPDLE	KGERLPQPPICITIDVYMIWVKCWM	960		
Db	901	HOSDVMWSYGVYVWELMT	FGAKPYDGI	PARBIPDLE	KGERLPQPPICITIDVYMIWVKCWM	960		
Qy	961	IDSECPRE	RELVSF	SRMARDP	QRFVW	IQNEDLGPASPLDSTFYRSLLED	DDMDGLVDA	1020
Db	961	IDSECPRE	RELVSF	SRMARDP	QRFVW	IQNEDLGPASPLDSTFYRSLLED	DDMDGLVDA	1020
Qy	1021	BEYLVPOQ	QFFCDDP	APAGAG	GMVHRRHS	SSSTRSGGDLTLGLSEPEE	EAAPRPLAPSEG	1080

Db	1021	EEYLVPOQFFCPDPAPGAGMWHRRSSSTRSGGGDLTLGLEPSEEEAPRSLAPSEG	1080
Qy	1081	AGSDVDFDGLMGAAKGLQLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVDFDGLMGAAKGLQLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKGVVVKOVFAFGGAVENPEYLTQP	1200
Db	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKGVVVKOVFAFGGAVENPEYLTQP	1200
Qy	1201	GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP	1255
RESULT 11			
AAE24067			
ID	AAE24067 standard; Protein; 1255 AA.		
XX	AC AAE24067;		
XX	23-SEP-2002 (first entry)		
XX	Human Her-2 protein.		
XX	Human; Her-2; epidermal growth factor receptor 2; infection; cancer;		
KW	hyperproliferative disorder; prophylaxis; inflammation; antisense;		
KW	tumour; gene therapy; phosphorothioate backbone.		
XX	Homo sapiens.		
OS	Homo sapiens.		
XX	WO200222636-A1.		
XX	21-MAR-2002.		
XX	12-SEP-2001; 2001WO-US28572.		
XX	15-SEP-2000; 2000US-0663834.		
XX	(ISIS-) ISIS PHARM INC.		
XX	Bennett CF, Cowseert LM;		
PI	WPI; 2002-471192/50.		
DR	N-PSDB; AAD38904.		
XX	Novel antisense oligonucleotide which modulates the expression of Human		
PT	Epidermal Growth Factor receptor, Her2, is useful for treating tumors		
PT	inflammation or to prevent infection in humans -		
XX	Example 13; Page 95-107; 116pp; English.		
XX	The invention relates to antisense compounds targetted to a nucleic		
CC	acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)		
CC	that specifically hybridises with and inhibits the expression of Her2.		
CC	Antisense compounds of the invention are used for treating diseases or		
CC	conditions associated with Her2 such as hyperproliferative disorders		
CC	e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,		
CC	neural or cardiac cancer. They are also useful prophylactically e.g.		
CC	to prevent or delay infection, inflammation and tumour formation. The		
CC	invention is also used in gene therapy. The present sequence is human		
CC	Her-2 protein.		
XX	Sequence 1255 AA;		
SQ			
Query Match 97.5%; Score 6631; DB 23; Length 1255;			
Best Local Similarity 97.6%; Pred. No. 0;			
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;			
Qy	1	MELAAICRWGLLALLPPGAASQVCTGTDMLKRLPASPETHDMLRHLVQGCVOGQNL	60
Db	1	MELAAICRWGLLALLPPGAASQVCTGTDMLKRLPASPETHDMLRHLVQGCVOGQNL	60

Qy	61	ELTYLPTNASLFLQDIEQVGYVLI AHNOVRQVPLQRLRIVRGTQQLFEDNYALAVLDNG	120
Db	61	ELTYLPTNASLFLQDIEQVGYVLI AHNOVRQVPLQRLRIVRGTQQLFEDNYALAVLDNG	120
Qy	121	DPLNTTPTVGTGSPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI FHKQNLQ	180
Db	121	DPLNTTPTVGTGSPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWKDI FHKQNLQ	180
Qy	181	LTLIDTNRSRACHPCSPMCKGSRGCMGSESSDCQSLTRITVCAGGCARCKGPLPTDCCHQC	240
Db	181	LTLIDTNRSRACHPCSPMCKGSRGCMGSESSDCQSLTRITVCAGGCARCKGPLPTDCCHQC	240
Qy	241	AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP	300
Db	241	AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP	300
Qy	301	YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCEKSKPCARVCYGLGHWQYI KANSKFIGIT	360
Db	301	YNYLSTDVGSCTLVCPHNOEVTAEQDTCRCEKSKPCARVCYGLGHWQYI KANSKFIGIT	360
Qy	361	ELPEFAGCKIFGSLAPLPESFDGDPASNTAPLQPEQLQVPELLEEITCYLYIISAWPDSLP	420
Db	361	IQEPFAGCKIFGSLAPLPESFDGDPASNTAPLQPEQLQVPELLEEITCYLYIISAWPDSLP	420
Qy	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLALIHNTLHLCFVHTV	480
Db	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLALIHNTLHLCFVHTV	480
Qy	481	PWDLFRQYIKANSKFIGITELECVGEGGLACHQICARGHCWGPGPTQCVCNSQFLRGQEC	540
Db	481	PWDLFRNPHQALLHTANRPEDECEVGEGLACHQICARGHCWGPGPTQCVCNSQFLRGQEC	540
Qy	541	VEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC	600
Db	541	VEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC	600
Qy	601	PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTISVAVVG	660
Db	601	PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAEQASPLTISVAVVG	660
Qy	661	ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL	720
Db	661	ILLVVLGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL	720
Qy	721	RKVKVLGSGAGFTYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVNAVGVSP	780
Db	721	RKVKVLGSGAGFTYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVNAVGVSP	780
Qy	781	YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRGLSQDGLLNWCHQIAKGMVLEDDR	840
Db	781	YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRGRGLSQDGLLNWCHQIAKGMVLEDDR	840
Qy	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDTETHADGKGVPIKWMALLESILRRFT	900
Db	841	LVRDLAARNVLKSPNHVKITDFGLARLLDIDTETHADGKGVPIKWMALLESILRRFT	900
Qy	901	HQSDVMSYGVTVVWELMTFGAKPDGIPARI PDLEKGERLPPOPICTIDVYIMLVKCMW	960
Db	901	HQSDVMSYGVTVVWELMTFGAKPDGIPARI PDLEKGERLPPOPICTIDVYIMLVKCMW	960
Qy	961	IDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA	1020
Db	961	IDSECRPRFRELVSFESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDGDLVDA	1020
Qy	1021	EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG	1080
Db	1021	EYLVPOQGFCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG	1080
Qy	1081	AGSDVDFDGLMGAAKGLQLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Db	1081	AGSDVDFDGLMGAAKGLQLPHTDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKGVVVKOVFAFGGAVENPEYLTQP	1200

Db 1141 NQDVRRPQPSREGPLPAARPAATLERPKTILSPGKNGVXKOVFAFGGAVENPEYLTPQ 1200  
1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPERGAPPSFKGTPTAENPEYLGLDVVP 1255  
1201 GGAAPQHPHPPAFSPAFDNLYYWDQPPERGAPPSFKGTPTAENPEYLGLDVVP 1255

RESULT 12  
AAE20479  
ID AAE20479 standard; Protein; 1255 AA.  
XX  
AC AAE20479;  
XX  
DT 01-JUL-2002 (first entry)  
XX  
DE Human Her-2/neu protein.  
XX  
KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 1021..1030  
FT /note= "Naturally processed HLA-B44-restricted epitope"  
XX  
PN W0200214503-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 14-AUG-2001; 2001WO-US41733.  
PR 14-AUG-2000; 2000US-225152P.  
PR 28-SEP-2000; 2000US-236428P.  
PR 21-FEB-2001; 2001US-270520P.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;  
PI McNeill PD, Vedvick TS;  
XX  
DR WPI; 2002-280758/32.  
DR N-PSDB; AAD32743.  
XX  
PT Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
PT prevention and diagnosis of cancer, preferably breast cancer -  
XX  
PS Disclosure; Page 114-117; 129pp; English.  
XX  
CC The invention relates to an isolated Her-2/Neu polypeptide composition  
CC effective for eliciting an immune response. The invention is useful for  
CC eliciting an immune response in a patient, where the patient is human  
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
CC The composition is useful for the therapy and diagnosis of cancer,  
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
CC and other compositions for the diagnosis, prevention and treatment of  
CC human malignancies, for stimulating and/or expanding T cells specific for  
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
CC patient. The invention is useful for stimulating a T cell response in a  
CC human patient, as probe or primer for nucleic acid hybridisation, to  
CC selectively form duplex molecules with complementary stretches of the  
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
CC length gene from a suitable library, and to direct expression of a  
CC polypeptide in appropriate host cells. The composition is useful in  
CC prophylactic or therapeutic applications and for the treatment of cancer,  
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
CC associated malignancies. The invention is useful in gene therapy. The  
CC present sequence is human Her-2/neu protein.  
XX  
SQ Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

Qy 1 MEAALCRWGLLALLPPGAASQTQCTGTDMKRLPAS PETHLDMLRHLYGCGVVOGNL 60  
Db 1 MEAALCRWGLLALLPPGAASQTQCTGTDMKRLPAS PETHLDMLRHLYGCGVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIJAHNOVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIJAHNOVROVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNNTPTVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNNOLA 180  
Db 121 DPLNNTPTVTGASPGGLREQLRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGSSEDCSLTRTVCGAGCARCKGPLTDCCHQEC 240  
Db 181 LTLIDTNRSRACHPCSPMCKSGRCWGSSEDCSLTRTVCGAGCARCKGPLTDCCHQEC 240

Qy 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMOYIKANSKFIT 360  
Db 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSKPCARVCYGLGMEHLREVRVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIYGLYISAWPSLP 420  
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETIYGLYISAWPSLP 420

Qy 421 DLSVFQNLQVIRGIRILHNGAYSLLTQGLGISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGIRILHNGAYSLLTQGLGISWLGRLSLRELGSGLALIHNTLHLCFVHTV 480

Qy 481 PWDQLFRQYIKANSKFITIGITELECVGEGLAACHQLCARGHCGWGPCTQCVNCSQFLRQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDCEVGEGLACHQLCARGHCGWGPCTQCVNCSQFLRQEC 540

Qy 541 VEECRVLQGLPREYVNRHCLPCHPECOPONGSVTCFGEADOCVACHYKDPFPCVARC 600  
Db 541 VEECRVLQGLPREYVNRHCLPCHPECOPONGSVTCFGEADOCVACHYKDPFPCVARC 600

Qy 601 PSGVKPDLSPYPIWKFPPDEBEGACQPCINCTHSCVDLDDKGCAPABQASPLTSTVSAVVG 660  
Db 601 PSGVKPDLSPYPIWKFPPDEBEGACQPCINCTHSCVDLDDKGCAPABQASPLTSTVSAVVG 660

Qy 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720

Qy 721 RKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780  
Db 721 RKVKVLGSGAFGTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780

Qy 781 YVSRLLGICLTSTVQLVTOLMPYGCCLLDHVRNRRGLSGODLLNWCQIAGKMSYLEDVR 840  
Db 781 YVSRLLGICLTSTVQLVTOLMPYGCCLLDHVRNRRGLSGODLLNWCQIAGKMSYLEDVR 840

Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWMALESILRRRFT 900

Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVWKCM 960

Qy 961 IDSECRPRELVSEFSRMRDQRFVWIONEDLGPASPLDSTFYSLLEDDDMGLVDA 1020  
Db 961 IDSECRPRELVSEFSRMRDQRFVWIONEDLGPASPLDSTFYSLLEDDDMGLVDA 1020

Qy 1021 EBYLVPQQGFFCPCDPAPGAGGMVHRHRSSTSRSGGDDLTGLSEPESEAPRPLAPSEG 1080  
Db 1021 EBYLVPQQGFFCPCDPAPGAGGMVHRHRSSTSRSGGDDLTGLSEPESEAPRPLAPSEG 1080

Db 1021 EYLVPOQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
 QY 1081 AGSDVFDGDLGMAKGLQSLPHDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
 Db 1081 AGSDVFDGDLGMAKGLQSLPHDPSPLQRYSDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
 QY 1141 NQDVRPQPPSPREGPLPAARPAAGATLRAKTLSPGKNGVVKOVFAFGGAVENPEYLTTPQ 1200  
 Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLRAKTLSPGKNGVVKOVFAFGGAVENPEYLTTPQ 1200  
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTTAENPEYLGLDVVP 1255  
 Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPSTFKGPTTAENPEYLGLDVVP 1255

## RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX PH Key Location/Qualifiers

XX FT Domain 1..653 /note= "extracellular domain"

XX FT Domain 676..1255 /note= "intracellular domain"

XX FT Domain 990..1255 /note= "phosphorylation domain"

XX PN WO200212341-A2.

XX PD 14-FEB-2002.

XX PF 03-AUG-2001; 2001WO-US24283.

XX PR 03-AUG-2000; 2000US-0632507.

XX PA (CORI-) CORIXA CORP.

XX PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Cheever MA, Gheysen D;

XX DR WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX PT Her-2/neu fusion protein for treating or preventing cancer by eliciting  
 or enhancing an immune response to the protein, has Her-2/neu  
 extracellular domain fused to Her-2/neu intracellular or  
 phosphorylation domain -  
 Claim 68; Fig 7; 141pp; English.

XX CC The present sequence is that of human Her-2/neu (p185 glycoprotein  
 or c-erbB2), an oncogenic self-protein and target for anti-cancer  
 vaccines. The Her-2/neu gene is amplified and p185 is overexpressed  
 in a variety of cancers, including breast, ovarian, colon, lung and  
 prostate cancer. Her-2/neu is a member of the tyrosine kinase  
 family of receptor-like glycoproteins. It comprises an extracellular  
 domain with homology to the epidermal growth factor receptor  
 (EGFR), a highly hydrophobic transmembrane domain and a C-terminal  
 intracellular domain that also shows homology to EGFR. Its  
 overexpression correlates with a poor prognosis in breast and  
 ovarian cancers. The invention provides Her-2/neu fusion  
 proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In  
 preferred fusion proteins, the extracellular domain of a Her-2/neu  
 protein is fused to a Her-2/neu intracellular domain or  
 phosphorylation domain (or its DeltaPD fragment). An immune  
 response to Her-2/neu protein is elicited or enhanced by  
 administering cells of an animal ex vivo with a vaccine, or by  
 transfecting cells of an animal in the form of a nucleic acid  
 encoding the fusion protein, and delivering the transfected cells  
 to the animal. The fusion proteins, nucleic acids, and isolated  
 specific T-cells are useful for inhibiting the development of a  
 cancer, especially breast, ovarian, colon, lung or prostate cancer  
 in a patient. T cells that specifically react with a Her-2/neu  
 fusion protein can be used to remove tumour cells from a sample in  
 order to inhibit the development of cancer in a patient.

XX SQ Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLVYGCQVQGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTQVCTGTDMLRLPASPEHLDMLRHLVYGCQVQGNL 60

QY 61 ELTYLPTNASLFLQDIOEVQGVYVIAHNOVROVPLQRLRVCTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLFLQDIOEVQGVYVIAHNOVROVPLQRLRVCTQLFEDNYALAVLDNG 120

QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDITLWKDIFHKNOLA 180

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPQLCYQDITLWKDIFHKNOLA 180

QY 181 LTLIDNRSRACHPCSPMCKGSRGWSESDCSQSLTRTVAGGCARCKGPIPTCCCHQC 240

Db 181 LTLIDNRSRACHPCSPMCKGSRGWSESDCSQSLTRTVAGGCARCKGPIPTCCCHQC 240

QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGCTLVCPPLHNOEVTAEQDTCRCEKSKPCARVCYGLGQYIKANSKFIGIT 360

Db 301 YNYLSTDVSGCTLVCPPLHNOEVTAEQDTCRCEKSKPCARVCYGLGMEHLREAVTSAN 360

QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETGLYLSAWPDSLP 420

Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEETGLYLSAWPDSLP 420

QY 421 DLSVFQNLQVIRGIRLHNGAYSLTLOGLGISWLGRLSLRELGLGSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGIRLHNGAYSLTLOGLGISWLGRLSLRELGLGSLALIHNNTHLCFVHTV 480

QY 481 PWDLFRQYIKANSKFITTELECVGSEGLACHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540

Db 481 PWDLFRQYIKANSKFITTELECVGSEGLACHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540

QY 541 VEESCRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

Db 541 VEESCRVLQGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600

QY 601 PSGVKPDLSPYMPITWKPDPDEGACQPCPINCTHSCVDLDDKGCPEQRASPLTSIVSAVVG 660

Db 601 PSGVKPDLSPYMPITWKPDPDEGACQPCPINCTHSCVDLDDKGCPEQRASPLTSIVSAVVG 660

QY 661 ILLVVLGVVFGILIKERQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRLIKETEL 720

Db 661 ILLVVLGVVFGILIKERQOKIRKYTMRLLOETELVEPLTPSGAMPNQAMRLIKETEL 720

QY 721 RKVKVLGSGAFGTYYGIIWIPDGENVKIPVAIKVLRNTPSKANKETLDBAYVMAGVSP 780

Db 721 RKVKVLGSGAFGTYYGIIWIPDGENVKIPVAIKVLRNTPSKANKETLDBAYVMAGVSP 780

QY 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDILLNWCQIAKMSYLEDVR 840  
DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDILLNWCQIAKMSYLEDVR 840  
QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900  
DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960  
QY 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVD 1020  
DB 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVD 1020  
QY 1021 EYLVPOQGFCCPDPAFGAGGMVHRHRSSTRSGGDLTLGLPESEEAAPRPLAPSEG 1080  
DB 1021 EYLVPOQGFCCPDPAFGAGGMVHRHRSSTRSGGDLTLGLPESEEAAPRPLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
QY 1141 NQDVPAPPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ 1200  
DB 1141 NQDVPAPPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1255  
DB 1201 GGAAPQHPHPPAFSPAFDNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1255

## RESULT 14

AAU77114  
ID AAU77114 standard; Protein; 1255 AA.

XX AC AAU77114;

XX DT 05-JUN-2002 (first entry)

XX DE Human Her-2/neu polypeptide.

XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;  
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;  
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;  
XX Hodgkin's lymphoma; T cell therapy.

XX OS Homo sapiens.

XX PN WO200213847-A2.

XX PD 21-FEB-2002.

XX PF 13-AUG-2001; 2001WO-US25408.

XX PR 14-AUG-2000; 2000US-0638280.

XX PR 28-SEP-2000; 2000US-0675904.

XX PA (CORI-) CORIXA CORP.

XX PI Gaiger A, Cheever MA, Hand-zimmermann S;

XX DR WPI; 2002-280741/32.

XX DR N-PSDB; ABX10730.

XX PT Inhibiting haematological malignancy development by administering  
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide  
PT encoding the polypeptide, or antigen presenting cells expressing the  
XX polypeptide  
XX PS Disclosure; Page 71-74; 74pp; English.

QY 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDILLNWCQIAKMSYLEDVR 840  
DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDILLNWCQIAKMSYLEDVR 840  
QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900  
DB 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVKCWM 960  
QY 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVD 1020  
DB 961 IDSECRPRFRELVSERFARMARDQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLVD 1020  
QY 1021 EYLVPOQGFCCPDPAFGAGGMVHRHRSSTRSGGDLTLGLPESEEAAPRPLAPSEG 1080  
DB 1021 EYLVPOQGFCCPDPAFGAGGMVHRHRSSTRSGGDLTLGLPESEEAAPRPLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
QY 1141 NQDVPAPPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ 1200  
DB 1141 NQDVPAPPPSPREGPLPAARPAAGATLERAKTILSPGKNGVYKDVFAFGGAVENPEYLTQ 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1255  
DB 1201 GGAAPQHPHPPAFSPAFDNLVYDQDPPERGAPPSTFKGTPTAENPEYLGLDV 1255

CC The invention relates to a method for inhibiting development of  
CC haematological malignancy in a patient by administering a polypeptide  
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
CC encoding the polypeptide. Antigen presenting cells that express the  
CC protein can also be administered. The sequences are used for inhibiting  
CC development of haematological malignancy such as acute myelogenous  
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
XX

QY Sequence 1255 AA;

Query Match 97.5%; Score 6631; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKURLPASPTHLDMLRHLVGGCCVQVGNL 60

DB 1 MELAALCRWGLLLALLPPGAASQVCTGTDMKURLPASPTHLDMLRHLVGGCCVQVGNL 60

QY 61 ELTYLPTNASLSFLQDIQEVGYVLAHNOVRQVPLQRLIRVRGTQLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASLSFLQDIQEVGYVLAHNOVRQVPLQRLIRVRGTQLFEDNYALAVLDNG 120

QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNOLA 180

DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNOLA 180

QY 181 LTLIDNTRSRACHPCSPMKGSRGWESSEDCOSLRTVTCAGGCARCKGLPTDCCHQC 240

DB 181 LTLIDNTRSRACHPCSPMKGSRGWESSEDCOSLRTVTCAGGCARCKGLPTDCCHQC 240

QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

QY 301 YNYLSTDVSGSCTLVCPHMQEVTAEDGTQCEKSKPCARVCYGLGHQYIKANSKFTGIT 360

DB 301 YNYLSTDVSGSCTLVCPHMQEVTAEDGTQCEKSKPCARVCYGLGHQYIKANSKFTGIT 360

QY 361 ELDFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVPEETLEETGYLYISAWPSLP 420

DB 361 IQBFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELOVPEETLEETGYLYISAWPSLP 420

QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLSLRSLRELGLSLALIHNNHLCFVHTV 480

DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLSLRSLRELGLSLALIHNNHLCFVHTV 480

QY 481 PWDQLPRQYIKANSKFTIGITELECVGEGLAHQLCARGHCWGPPTQCVNCSQFLRQEC 540

DB 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPPTQCVNCSQFLRQEC 540

QY 541 VEECRVLQGLPREYVYVNAHCLPCHPECQONGSVTCFGEADQCVCAHYKDPFPCVARC 600

DB 541 VEECRVLQGLPREYVYVNAHCLPCHPECQONGSVTCFGEADQCVCAHYKDPFPCVARC 600

QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPEARASPLTISVSAVVG 660

DB 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPEARASPLTISVSAVVG 660

QY 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

DB 661 ILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720

QY 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780

DB 721 RKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKIILDEAYMAGVGP 780

QY 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDILLNWCQIAKMSYLEDVR 840

DB 781 YVSRLLGICLTSTVQLVQLMPYGCCLLDHVRNRRGRGLSGDILLNWCQIAKMSYLEDVR 840

QY 841 LVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFT 900

|||||  
841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRFT 900  
QY  
901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCM 960  
DB  
901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCM 960  
QY  
961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1020  
DB  
961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1020  
QY  
1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTGLLEPSEEAAPRSLAPSEG 1080  
DB  
1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTGLLEPSEEAAPRSLAPSEG 1080  
QY  
1081 AGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
DB  
1081 AGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY  
1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYLT PQ 1200  
DB  
1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLPSPGKNGVVKDVFAFGGAVENPEYLT PQ 1200  
QY  
1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVVP 1255  
DB  
1201 GGAAPQHPHPPAFSPAFDNLVYWDQPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

## RESULT 15

AAR39568  
ID AAR39568 standard; Protein; 1433 AA.XX  
AC AAR39568;XX  
DT 07-FEB-1994 (first entry)XX  
DE Sequence of c-erbB-2 tumour antigen.XX  
KW Tumour antigen; c-erbB-2; glycoprotein.XX  
OS Homo sapiens.XX  
PN W09316185-A.XX  
PD 19-AUG-1993.XX  
PF 05-FEB-1993; 93WO-US01055.XX  
PR 06-FEB-1992; 92US-0831967.XX  
PA (CETU ) CETUS ONCOLOGY CORP.XX  
PI (CREA-) CREATIVE BIOMOLECULES INC.XX  
PI Houston LL, Huston JS, Oppermann H, Ring DB;XX  
DR WPI; 1993-272889/34.XX  
DR N-PSDB; AAQ46083.XX  
PT New single chain Fv polypeptide binding to C-erbB-2 tumourXX  
PT antigen - for imaging or treating breast or ovarian cancer etc.XX  
PS Disclosure; pages 48-54; 87pp; English.XX  
CC c-erbB-2 refers to a protein antigen expressed on the surface ofXX  
CC tumour cells, such as breast and ovarian tumour cells, which is anXX  
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectricXX  
CC pt. of about 5.3 (see AAQ46083; AAR39568). X in AAR39568 representsXX  
CC the location of a stop codon in AAQ46083.XX  
SQ Sequence 1433 AA;

Query Match

Best Local Similarity

96.8%; Score 6588; DB 14; Length 1433;

97.1%; Pred. No. 0;

Matches 1218; Conservative 8; Mismatches 29; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLKRLPASPETHLDMRLHLYQCGVQVQGNL 60  
DB 1 MELAALCRWGLLLALLPPGAASTOVCTGTDMLKRLPASPETHLDMRLHLYQCGVQVQGNL 60  
QY 61 ELTYLPTNASISFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNYALAVL DNG 120  
DB 61 ELTYLPTNASISFLQDIOEVQGYVLI AHNQVRQVPLQRLRIVRGTLQFEDNYALAVL DNG 120  
QY 121 DPLNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180  
DB 121 DPLNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180  
QY 181 LTLTDVNRSRACHPCSPMKGSRGWSSESSEDCQSLTRTVCAAGGACRCKGPLPTDCCHQEC 240  
DB 181 LTLTDVNRSRACHPCSPMKGSRGWSSESSEDCQSLTRTVCAAGGACRCKGPLPTDCCHQEC 240  
QY 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIOIT 360  
DB 301 YNYLSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360  
QY 361 ELEFAGCKIFGSLAFLPESFDGDPASNTAPLQPEOLQVFFETLEETGTYLISAWPDSL 420  
DB 361 IOEPAGCKRIFGSLAFLPESFDGDPASNTAPLQPEHLQVFFETLEETGTYLISAWPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLISWLGSLRLSRLGSLALIHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLISWLGSLRLSRLGSLALIHNNTHLSFVHTV 480  
QY 481 PWDQLFRQYIKANSKFIOITTELECVGEGGLACHQICARGHCWGPGTQCWCNSQFLRGQEC 540  
DB 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCWGPGTQCWCNSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPECVAVC 600  
DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPPECVAVC 600  
QY 601 PSYKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPACPAEQASPLTSIVAVG 660  
DB 601 PSYKVPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPACPAEQASPLTSIVAVG 660  
QY 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQWIRILKETEL 720  
DB 661 ILLVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQWIRILKETEL 720  
QY 721 RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
DB 721 RKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
QY 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNCMQIAKMSYLEVDV 840  
DB 781 YVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSGDQLLNCMQIAKMSYLEVDV 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRFT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRRFT 900  
QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCM 960  
DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKCM 960  
QY 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1020  
DB 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDDDMDGLVDA 1020  
QY 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTGLLEPSEEAAPRSLAPSEG 1080  
DB 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTGLLEPSEEAAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVFPQPPSPREGPLPAARPAAGATLERAKTLPSPGKNGVVKDVFAGGAVENPEYLTPO	1200
Db	1141	NQPDVFPQPPSPREGPLPAARPAAGATLERPKTLPSPGKNGVVKDVFAGGAVENPEYLTPO	1200
Qy	1201	GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPAENPEYLGLDVTV	1255
Db	1201	GGAAPQPHPPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPAENPEYLGLDVTV	1255

Search completed: July 22, 2003, 08:41:34  
 Job time : 42.9774 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds  
 (without alignments)  
 5088.033 Million cell updates/sec

Title: SEQ4-465-479-12  
 Perfect score: 6804  
 Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6631	97.5	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5851.5	86.0	1254	1 ERB2 MESAU	P06553 mesocricetu
3	5850	86.0	1257	1 ERB2 RAT	P06494 rattus norv
4	3119	45.8	1210	1 EGFR HUMAN	P00533 homo sapien
5	3088	45.4	1210	1 EGFR MOUSE	Q01279 mus musculu
6	2943.5	43.3	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2921	42.9	1308	1 ERB4 RAT	Q62956 rattus norv
8	2674.5	39.3	1167	1 XMRK XIPMA	P13388 xiphophoru
9	2393.5	35.2	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2323.5	34.1	1339	1 ERB3 RAT	Q62799 rattus norv
11	1925	28.3	1426	1 EGFR DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB AVIER	P00535 avian eryth
14	1630	24.0	540	1 ERBB AVIEU	P11273 avian eryth
15	1575	23.1	703	1 EGFR CHICK	P13387 gallus gall
16	1292	19.0	1223	1 LTK3 CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2 MOUSE	P70424 mus musculu
18	724	10.6	1363	1 ILPR BRALA	O02466 brachioosto
19	696	10.2	1300	1 IRR MOUSE	Q9wt14 mus musculu
20	695	10.2	1383	1 INSR RAT	P15127 rattus norv
21	695	10.2	1477	1 HTK7 HYDAT	Q25197 hydra atten
22	694.5	10.2	1372	1 INSR MOUSE	Q25198 mus musculu
23	694	10.2	1382	1 INSR HUMAN	P06213 homo sapien
24	691	10.2	1297	1 IRR HUMAN	P14616 homo sapien
25	685	10.1	1607	1 MIPR LYMTS	Q25410 lymnaea sta
26	682.5	10.0	1300	1 IRR CAVPO	P14617 cavia porce
27	646	9.5	1367	1 IGR1 HUMAN	P08069 homo sapien
28	629	9.2	1373	1 IGR1 MOUSE	Q60751 mus musculu
29	626	9.2	1390	1 INSR AEDAE	Q93105 aedes aegypt
30	625.5	9.2	1370	1 IGR1 RAT	P24062 rattus norv
31	616	9.1	2146	1 INSR DROME	P09208 drosophila
32	609.5	9.0	984	1 EPB1 RAT	P09759 rattus norv
33	603.5	8.9	984	1 EPB1 HUMAN	P54762 homo sapien

RESULT 1  
 ERB2\_HUMAN  
 ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
 AC P04626;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 13-AUG-1987 (Rel. 05, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).  
 DE ERBB2 OR HER2 OR NGL OR NEU.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86118663; PubMed=3003577;  
 RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;  
 RT "Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth factor receptor.";  
 RL Nature 319:230-234(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86070181; PubMed=2999974;  
 RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Francke U., Levinson A., Ullrich A.;  
 RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene.";  
 RL Science 230:1132-1139(1985).  
 RN [3]  
 RP SEQUENCE OF 737-1031 FROM N.A.  
 RX MEDLINE=86016729; PubMed=2995967;  
 RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
 RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
 RN [4]  
 RP VARIANTS VAL-654 AND VAL-655.  
 RX MEDLINE=93194196; PubMed=8095488;  
 RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
 RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization.";  
 RL Genomics 15:426-429(1993).  
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

34 603 8.9 987 1 EPB4\_HUMAN P54760 homo sapien  
 35 601.5 8.8 984 1 EPB1\_CHICK Q07494 gallus gall  
 36 593 8.7 902 1 EPBB\_XENLA Q91736 xenopus lae  
 37 587.5 8.6 985 1 EPBA\_XENLA Q91571 xenopus lae  
 38 587 8.6 1114 1 RET\_HUMAN P07949 homo sapien  
 39 585.5 8.6 977 1 EPA2\_MOUSE Q03145 mus musculu  
 40 585 8.6 987 1 EPB4\_MOUSE P54761 mus musculu  
 41 584.5 8.6 976 1 EPA2\_HUMAN P29317 homo sapien  
 42 578.5 8.5 1053 1 FAK1\_CHICK Q00944 gallus gall  
 43 573 8.4 757 1 HT16\_HYDAT P53356 hydra atten  
 44 569 8.4 1068 1 FAK1\_XENLA Q91738 xenopus lae  
 45 565.5 8.3 1052 1 FAK1\_MOUSE P34152 mus musculu

## ALIGNMENTS



CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -!- SIMILARITY: BELONGS TO THE EGFR RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M11767; AAA35808.1; -;  
 CC EMBL; M11761; AAA35808.1; JOINED.  
 CC EMBL; M11762; AAA35808.1; JOINED.  
 CC EMBL; M11763; AAA35808.1; JOINED.  
 CC EMBL; M11764; AAA35808.1; JOINED.  
 CC EMBL; M11765; AAA35808.1; JOINED.  
 CC EMBL; M11766; AAA35808.1; JOINED.  
 CC EMBL; M11730; AAA7493.1; -;  
 CC EMBL; M12036; AAA35978.1; -;  
 CC EMBL; X03363; CAA27060.1; -;  
 CC PIR; A25491; A25491.  
 CC PIR; A24571; A24571.  
 CC HSP; P11362; 1FGK.  
 CC Genew; HGNC:3430; ERBB2.  
 CC MIM; 164870; -;  
 CC InterPro; IPR000494; EGFR\_L\_domain.  
 CC InterPro; IPR000719; Euk\_pkinase.  
 CC InterPro; IPR002174; Furin-like.  
 CC InterPro; IPR001245; Tyr\_pkinase.  
 CC InterPro; IPR004019; YLP\_motif.  
 CC Pfam; PF00069; pkinase; 1.  
 CC Pfam; PF00757; Furin-like; 1.  
 CC Pfam; PF01030; Recep\_L\_domain; 2.  
 CC Pfam; PF02757; YLP; 2.  
 CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC SMART; SM00361; FU; 3.  
 CC SMART; SM00319; TyxKc; 1.  
 CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 CC PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Polymorphism.  
 CC SIGNAL 1 21 POTENTIAL.  
 CC CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 CC DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 653 675 POTENTIAL.  
 CC DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 720 987 PROTEIN KINASE.  
 CC NP\_BIND 726 734 ATP (BY SIMILARITY).  
 CC BINDING 753 753 ATP (BY SIMILARITY).  
 CC ACT\_SITE 845 845 BY SIMILARITY.  
 CC DISULFID 195 204 BY SIMILARITY.  
 CC DISULFID 199 212 BY SIMILARITY.  
 CC DISULFID 220 227 BY SIMILARITY.  
 CC DISULFID 224 235 BY SIMILARITY.  
 CC DISULFID 236 244 BY SIMILARITY.  
 CC DISULFID 240 252 BY SIMILARITY.  
 CC DISULFID 255 264 BY SIMILARITY.  
 CC DISULFID 268 295 BY SIMILARITY.  
 CC DISULFID 299 311 BY SIMILARITY.  
 CC DISULFID 315 331 BY SIMILARITY.  
 CC DISULFID 334 338 BY SIMILARITY.  
 CC DISULFID 511 520 BY SIMILARITY.  
 CC DISULFID 515 528 BY SIMILARITY.  
 CC DISULFID 531 540 BY SIMILARITY.  
 CC DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC) (POTENTIAL).  
 FT VARIANT 654 654 I -> V.  
 FT VARIANT 655 655 /FTID=VAR\_004077.  
 FT VARIANT 1170 1170 I -> V.  
 FT CONFLICT 1255 AA; 137909 MW; 39E9DFA04DCF962 CRC64;  
 FT SEQUENCE  
 SQ  
 Query Match 97.5%; Score 6631; DB 1; Length 1255;  
 Best Local Similarity 97.6%; Pred. No. 0;  
 Matches 1225; Conservative 6; Mismatches 24; Indels 0; Gaps 0;  
 Qy 1 MELAALCRWGLLLALLPFGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVVQGNL 60  
 Db 1 MELAALCRWGLLLALLPFGAASQVCTGTDMKRLPASPTHLDMLRHLVGGCVVQGNL 60  
 Qy 61 ELTYLPTNASLSFLQDIQEVQVYLI AHNOVRQVPLQRLIRVRCGTOLPFDNYALAVLNG 120  
 Db 61 ELTYLPTNASLSFLQDIQEVQVYLI AHNOVRQVPLQRLIRVRCGTOLPFDNYALAVLNG 120  
 Qy 121 DPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180  
 Db 121 DPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180  
 Qy 181 LTLIDTNRSRACHPCSPMKGSRGWGSESSDCSLTRTVCGAGCARCKGRLPTCCCHQC 240  
 Db 181 LTLIDTNRSRACHPCSPMKGSRGWGSESSDCSLTRTVCGAGCARCKGRLPTCCCHQC 240  
 Qy 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Qy 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGHQYIKANSKFIGIT 360  
 Db 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGHQYIKANSKFIGIT 360  
 Qy 361 ELSEFAGCKTFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEETIGLYLISAWPDSL 420  
 Db 361 IQSEFAGCKTFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEETIGLYLISAWPDSL 420  
 Qy 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRSLRGLSLALIHNNTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLGISWGLRSLRSLRGLSLALIHNNTHLCFVHTV 480  
 Qy 481 PWDQLFRQYIKANSKFITIGITELECVGEGGLACHOLCARGHCWGPGPTQCVNCSQFLRGQC 540  
 Db 481 PWDQLFRQYIKANSKFITIGITELECVGEGGLACHOLCARGHCWGPGPTQCVNCSQFLRGQC 540  
 Qy 541 VEESCRVLQGLPREYVNAHCLPCHPECPONGSVTCFPEADQCVACAHKDPDFCVARC 600  
 Db 541 VEESCRVLQGLPREYVNAHCLPCHPECPONGSVTCFPEADQCVACAHKDPDFCVARC 600  
 Qy 601 PSGVKPDLSTYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSTVSAVVG 660  
 Db 601 PSGVKPDLSTYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCPEAQASPLTSTVSAVVG 660  
 Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720

```
Qy 721 RKVVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Db 721 RKVVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780
Qy 781 YVSRLLGICITSTVOLVTLQMPYGCCLLDHVRENRGLQSDLLNWCQIAKGMYSLEDDR 840
Db 781 YVSRLLGICITSTVOLVTLQMPYGCCLLDHVRENRGLQSDLLNWCQIAKGMYSLEDDR 840
Qy 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALESILRRFT 900
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960
Qy 961 IDSECRFRFRELVSFMRMARDPQFVVVQIENEDLGASPLDSTFYRSLLEDGMDLVA 1020
Db 961 IDSECRFRFRELVSFMRMARDPQFVVVQIENEDLGASPLDSTFYRSLLEDGMDLVA 1020
Qy 1021 EYLVLPQGFCCPDPAFAGAGMVRHRSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Db 1021 EYLVLPQGFCCPDPAFAGAGMVRHRSSTRSGGDLTLGLPSEBEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NQDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Db 1141 NQDVRPQPPSPREGPLPAARPAAGATLAKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200
Qy 1201 GGAAPQHPHPPAFSPADNLVYWDQPPERGAPPSTFGTPTANPEYLGLDVVP 1255
Db 1201 GGAAPQHPHPPAFSPADNLVYWDQPPERGAPPSTFGTPTANPEYLGLDVVP 1255

RESULT 2
ERB2 MESAU
ID ERB2 MESAU STANDARD; PRT; 1254 AA.
AC Q6053;
DT 15-DEC-1998 (Rel. 37, Last Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 13-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GS ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GF30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
```

```
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D16295; BAA03801.1; -.
CC HSP; P11362; IFGK.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC InterPro; IPR004019; YLP motif.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC Pfam; PF02757; YLP; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Proto-oncogene; Disease mutation.
CC SIGNAL 1 21 POTENTIAL.
CC CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
CC DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 653 675 POTENTIAL.
CC DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).
CC DOMAIN 158 368 CYS-RICH.
CC DOMAIN 472 644 CYS-RICH.
CC DOMAIN 720 987 PROTEIN KINASE.
CC NP_BIND 726 734 ATP (BY SIMILARITY).
CC BINDING 753 753 ATP (BY SIMILARITY).
CC ACT_SITE 845 845 BY SIMILARITY.
CC DISULFID 195 204 BY SIMILARITY.
CC DISULFID 199 212 BY SIMILARITY.
CC DISULFID 236 244 BY SIMILARITY.
CC DISULFID 240 252 BY SIMILARITY.
CC DISULFID 255 264 BY SIMILARITY.
CC DISULFID 268 295 BY SIMILARITY.
CC DISULFID 299 311 BY SIMILARITY.
CC DISULFID 315 331 BY SIMILARITY.
CC DISULFID 334 338 BY SIMILARITY.
CC DISULFID 511 520 BY SIMILARITY.
CC DISULFID 515 528 BY SIMILARITY.
CC DISULFID 531 540 BY SIMILARITY.
CC DISULFID 544 560 BY SIMILARITY.
CC DISULFID 563 576 BY SIMILARITY.
CC DISULFID 567 584 BY SIMILARITY.
CC DISULFID 587 596 BY SIMILARITY.
CC DISULFID 600 623 BY SIMILARITY.
CC DISULFID 626 634 BY SIMILARITY.
CC DISULFID 630 642 BY SIMILARITY.
CC MOD_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC MOD_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC CARBOHYD 68 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 125 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 187 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 259 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 530 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 571 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 571 N-LINKED (GLCNAC...) (POTENTIAL).
CC CARBOHYD 629 V -> E (IN ONCOGENIC NEU).
CC CARBOHYD 658 V -> E (IN ONCOGENIC NEU).
CC VARIANT 659 659 V -> E (IN ONCOGENIC NEU).
CC SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;
Query Match 86.0%; Score 5851.5; DB 1; Length 1254;
```



```

or send an email to licensee@sib-sib.ch).
-----
CC ENBL; X03362; CAA27059.1; ALT_INIT.
DR PIR; A24562; TVRTNU.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR DR pfam; PF00069; pkinase; 1.
DR DR pfam; PF00757; Furin-like; 1.
DR DR pfam; PF01030; Recep_L_domain; 2.
DR DR pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKC; 1.
DR PROSITE; PS0107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.
FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 655 677 POTENTIAL.
FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 159 369 CYS-RICH.
FT DOMAIN 473 646 CYS-RICH.
FT DOMAIN 722 989 PROTEIN KINASE.
FT NP_BIND 728 736 ATP (BY SIMILARITY).
FT BINDING 755 755 ATP (BY SIMILARITY).
FT ACT_SITE 847 847 BY SIMILARITY.
FT DISULFID 196 205 BY SIMILARITY.
FT DISULFID 200 213 BY SIMILARITY.
FT DISULFID 221 228 BY SIMILARITY.
FT DISULFID 225 236 BY SIMILARITY.
FT DISULFID 237 245 BY SIMILARITY.
FT DISULFID 241 253 BY SIMILARITY.
FT DISULFID 256 265 BY SIMILARITY.
FT DISULFID 269 296 BY SIMILARITY.
FT DISULFID 300 312 BY SIMILARITY.
FT DISULFID 316 332 BY SIMILARITY.
FT DISULFID 335 339 BY SIMILARITY.
FT DISULFID 513 522 BY SIMILARITY.
FT DISULFID 517 530 BY SIMILARITY.
FT DISULFID 533 542 BY SIMILARITY.
FT DISULFID 546 562 BY SIMILARITY.
FT DISULFID 565 578 BY SIMILARITY.
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
Query Match 86.0%; Score 5850; DB 1; Length 1257;
Best Local Similarity 86.2%; Pred.No. 3.4e-305;
Matches 1084; Conservative 51; Mismatches 120; Indels 2; Gaps 2;

```

```

QY 1 MELAAALCRWGLLLALLPPGAASCTGTGDMKLRLPASPETHLDMLRLHLYQCQVVOGNL 60
DB 1 MELAAACRWGLFALLPPGIAGTCTGTGDMKLRLPASPETHLDMLRLHLYQCQVVOGNL 60

```

```

QY 61 ELTYLPTNASLFLQDIOEQVGVYVLIHAHNOVROVPLQRLRIVRGTOQLFEDNYALAVLDNG 120
DB 61 ELTYVVPANASLSFLQDIOEQVGVYVLIHAHNOVKRVPQLRLRIVRGTOQLFEDKYALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGRLRELQLRSLTEILKGGVLIORNPOLCYQDTILWKDIFHNNQL 179
DB 121 DPQDNVAASPGRTPGRLRELQLRSLTEILKGGVLIIRGNPOLCYQDMVLMKDFVRKNQL 180
QY 180 ALTLIDITNRSRACHPCSPMCKSCRCGSESSCOSLTRTVCAGGCACRCKGLPTDCCHEQ 239
DB 181 APVDIDITNRSRACPPCAPACKDNHCWSESPEDCOILTGTICTSCACRCKGLPTDCCHEQ 240
QY 240 CAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMNPGRYTFGASCVTTC 300
QY 300 PNYLSTDVGSCTLVCPHLNHNQVTAEDGTORCKSKPCARVCYGLGMQYIKANSKFTGI 359
DB 301 PNYLSTEVGSCTLVCPNPNQVTAEDGTORCKSKPCARVCYGLGMHRLGARAITSD 360
QY 360 TELEFAGCKKI FGSLLAFLPESPDGDPASNTAPLOPELOVPFETLEEITGVLYISAWPDSL 419
DB 361 NVQEFQCKKIFGSLLAFLPESPDGDPSSGIAPLRPEQLQVFETLEEITGVLYISAWPDSL 420
QY 420 PDLVSFQNLQVIRGRIILHNGAYSLTLQGLGISWLGRLSLRELGLALIHNNTHLCFVHT 479
DB 421 RDLVSFQNLRIIRGRIILHNGAYSLTLQGLGISWLGRLSLRELGLALIHNNTHLCFVHT 480
QY 480 VPHDQLFROYIKANSKFTIGITELE-CVGEGLACHQICARHCWGPGTQCNCQSFURGQ 538
DB 481 VPMQDLFRNPHQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCGPGTQCNCQSHFLRGQ 540
QY 539 ECVEECRVLQGLPREVYNARHCLPCHPEQOPONGSVTCFGEADOCVACAHYKDPFCVA 598
DB 541 ECVEECRVKGLPREYVSDKRLPCHPEQOPONSSETCFGEADQCAACAHYKDSSCVA 600
QY 599 RCPGSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPQORASPLTISVAV 658
DB 601 RCPGSGVKPDLSYMPIWKYPDEEGI CQPCPINCTHSCVDLDERGCPAQORASPTVFIATV 660
QY 659 VGIILLVVVGVVPGIILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNOAQRILKET 718
DB 661 VGVLLFLILVVVVGILIKRRRQKIRKYTWRRLLQETELVEPLTPSGAMPNOAQRILKET 720
QY 719 ELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAVG 778
DB 721 ELRKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAVG 780
QY 779 SPVSRLLGICLTSTVQLVTQMPYGCLLDHDVRENRGLSGQDLLNMCMOIAKGMSTYLED 838
DB 781 SPVSRLLGICLTSTVQLVTQMPYGCLLDHDVRENRGLSGQDLLNMCVQIAKGMSTYLED 840
QY 839 VRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRR 898
DB 841 VRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRRR 900
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMWKVC 958
DB 901 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMWKVC 960
QY 959 WMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDDDMGDLV 1018
DB 961 WMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPSSPMDSTFYRSLLEDDDDMGDLV 1020
QY 1019 DABEYLVPOQGFCDPDPAPGAGMVHRRHSSTRSGGDLTLGLPSEEEAPRSLAPS 1078
DB 1021 DABEYLVPOQGFSPDPTPGTGSTAHRHRRSSSTRSGGDLTLGLPSEEGPSPSLAPS 1080
QY 1079 EGAGSDVDFDGLGMGAAGKLSLPHDPSPLORYSEDPTVPLPSETDGVVAPLTCSPQPE 1138
DB 1081 EGAGSDVDFDGLAMGVTKGLQSLSPHDLSPLOYSDEPTLPLPPTDGVVAPLACSPQPE 1140

```

Qy 1139 YVNPQDVRRPPSPREGPLPAARAGATLERAKTLSPGKNGVWVDVFAFGAVENPEYL 1198  
 Db 1141 YVNGSEVQPQPLTPEGPLPVRAGATLERPKTLSPGKNGVWVDVFAFGAVENPEYL 1200  
 Qy 1199 PQGGAAPQHPHPAPAFSADFNLVYWDQPPRGERAPPSTFKTPTAENPEYLGLDVVP 1255  
 Db 1201 PREGTAPHPSPAFSADFNLVYWDQNSSEQGPPSNFECTPTAENPEYLGLDVVP 1257

RESULT 4  
 EGRF\_HUMAN  
 ID EGRF\_HUMAN STANDARD; PRT; 1210 AA.  
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;  
 AC O00686; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 DE protein-tyrosine kinase ErbB-1).  
 OS Homo sapiens (Human).  
 GN EGRF OR ERBB1.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RA "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RL Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=95382957; PubMed=7653368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RA "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reiter J.L., Maible N.J.;  
 RA "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;  
 RA "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RC TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Maible N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RL isoforms.";  
 RN Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

MAIBLE N.J.;  
 "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RA "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RL cells.";  
 RN Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RA "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RA "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RA "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RA "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RL factor receptor proto-oncogene RNA synthesis.";  
 RN J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RA "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RX Kohda D.;  
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Morig G., Cohen S.;  
 RA "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RA "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RL Identification of a novel site in EGF receptor.";  
 RN J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]

RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 the epidermal growth factor receptor expressed in Chinese hamster  
 ovary fibroblasts.";  
 RL Growth Factors 13:121-132 (1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731688;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 An-X-Cys sequence of recombinant human epidermal growth factor  
 receptor.";  
 RL J. Biochem. 127:65-72 (2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Odaka M., Inagaki F., Lax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor.";  
 RL J. Biol. Chem. 273:11150-11157 (1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 mitogens.";  
 RL Annu. Rev. Biochem. 56:881-914 (1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 EGF-like growth factor, Gp30 and vaccinia virus growth factor. Is  
 involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 dimerization, internalization of the EGF-receptor complex,  
 induction of the tyrosine kinase activity, stimulation of cell  
 synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X00588; CAA25240.1; -  
 DR EMBL; U95083; AAB53063.1; -  
 DR EMBL; U48722; AAC50802.1; -  
 DR EMBL; U48723; AAC50804.1; -  
 DR EMBL; U48724; AAC50796.1; -  
 DR EMBL; U48725; AAC50797.1; -  
 DR EMBL; U48726; AAC50798.1; -  
 CC  
 Query Match 45.8%; Score 3119; DB 1; Length 1210;  
 Best Local Similarity 49.7%; Pred. No. 2.4e-159;  
 Matches 629; Conservative 173; Mismatches 356; Indels 108; Gaps 23;  
 QY 11 LLLALLPPGAA--STOVCTGTDMLRLPASPETHLDMLRHLHYOGCVVQGNLTYLPTN 68  
 DB 11 LLLALLPPGAA--STOVCTGTDMLRLPASPETHLDMLRHLHYOGCVVQGNLTYLPTN 68  
 QY 14 LLAALCPASRALEBKVCQGSNKLTLQLTGTFDFHFLSLQRMFNNCVVLGNLITYVQRN 73  
 DB 14 LLAALCPASRALEBKVCQGSNKLTLQLTGTFDFHFLSLQRMFNNCVVLGNLITYVQRN 73

QY 69 ASLSFLDIOEVQGVYLIAHNQVRQVPLQRLRIVRGTQLPEDNYALAVLDGDPDLNLTTP 128  
 DB 74 YDLSFLTKIOEVAGYVLIALTNTVERIPLENQIIRGNMYENSVALAVLSNYD----- 126  
 QY 129 VTGASPGGLRELQLRSITELKGGVLIORNPQLCYQDTILWKDIFPHKNQALALTLIDTNR 188  
 DB 127 ---ANKTGLKELPMRNLEILHGAVRFSNNPALCNVESIQWRDIVSSDFLSNMSDFQNH 183  
 QY 189 SRACHPCSPMKGSCWGESSEDCOSLTRTVACGGCA-RCKGPLPTDCHCEQACGCTGP 247  
 DB 184 LGSCQKCDPSCPNGSCWGEENCOKTKIICAQCSGCRCKGSPSCDCHNQCAAGCTGP 243  
 QY 248 KHSCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPYNVLTSD 307  
 DB 244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYQMDVNPPEGKYSFGATCKKCPRYVVD 303  
 QY 308 VGSCTLVCPLNHOEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGITELE-FAG 366  
 DB 304 HGSCVTRACGADSYEM-BEDGVRKCKCEGPCRVKCVNGIGIEFK-DSLSINATNIKHPKN 361  
 QY 367 CKKIFGSLAFLPESFDGDPASNTAPLOEQLOVPETLEEITGYLYISAWPDSLPLDSVQ 426  
 DB 362 CTSISGDLHILPVAFRGDSFTHTPPLDPQELDLIKTVEITGFLLIQAWPENRTDLHAPE 421  
 QY 427 NLQVIRGIRLHNGAYSLTLOGLGISWGLSLRSLRELGLALIHHTHLFCFVHTVPWOLF 486  
 DB 422 NLEIRGRTKHQGFSLAVVSLNITSGLSLRSLKELISDGDVLIISGNKLYCANTINWKLFL 481  
 QY 487 RQYIKANSKFIGITELECVGEGELACHOLCARGHCWGPGTQCVNCSOFLROECVEECRV 546  
 DB 482 GTSQKTKIISNRGENSKATQGVCHALCSPEGCWGPEPRDCVSCRVNVRGECVDCNVL 541  
 QY 547 LQGLPREYVNAHCLPCHPSCQNGSVTCFPGPADOCVACAHYKDPFCVAPCSGVKP 606  
 DB 542 LEGERPREFVENSECTIQCHPELPOAMNITCTGRGPDNCIQCAHYIDGPHCVKTCPCAGVMG 601  
 QY 607 DLSYMPIWKFPEDEGACQPCPCINCTHSCVDLDDKCPAEQASPLTSTVSVAVG--ILL 663  
 DB 602 ENNTL-VKRYADAGHVCHLCHPCNTYCTGFGLEGCTNGPKIP-SIATGVGALLULL 658  
 QY 664 VVVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNQOMRILKETELRV 723  
 DB 659 VVALGIG---LFMRRRHIVRKRLLRLLQERELVEPLTPSGEAPQALLRLKETEPKKI 715  
 QY 724 KVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVSPVVS 783  
 DB 716 KVLGSGAGTVYKGLWIPGEKVKIPVAIKELREATSPKANKEILDEAYVNASVDNPHVC 775  
 QY 784 RLLGICLTSTVQLVTQMLPYGCLLDHVRENKRLGSODLLNWCMIKAGMSYLEDVRLVH 843  
 DB 776 RLLGICLTSTVQLITQMLPFGLLDYVREKDNIGSQVLLNWCMIKAGMYLEDVRLVH 835  
 QY 844 RDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILRRFFTHQS 903  
 DB 836 RDLAARNVLVKTPQHVKITDFGLAKLGAEEKVEYHSGGKVPKIKWMALESILHRYTHQS 895  
 QY 904 DWMSYGVTVWELMTFGNKPVDGIPAREIPDLLEKGERLPQPPICTIDVYMIWVCWMIDS 963  
 DB 896 DWMSYGVTVWELMTFGSKPYDGIIPASEISILEKGERLPQPPICTIDVYMIWVCWMIDA 955  
 QY 964 ECRPRFRELVSFERNARDPQRFVIO-NEDLGPASPLDSTFYRSLLEDDDDMGDLVDAEE 1022  
 DB 956 DSRPKFRELIIESKWARDPQRYLVIOGDERMLPSTDSNFYRALMDEEDDDMDVDVADE 1015  
 QY 1023 YLVPQGGFFCPDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEBEAPRSLAPSEAG 1082  
 DB 1016 YLVPQGGFF-----SSPSTRTPPLLSLSAT 1041  
 QY 1083 SDVFDGDLGMAAKGLOSPLTHDPSPQRYSEDPVLPSET--DGVPAPLTCSPQPEYV 1140  
 DB 1042 SN--NSTVACIDRNLQSCPIKEDSFLQRYSSDPTGALTEDSDIDDTFL-----PVPEYI 1093





```

FT MOD_RES 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
      (BY SIMILARITY).
FT CARBOHYD 128 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 C -> S (IN REF. 2).
FT CONFLICT 539 C -> W (IN REF. 5).
FT CONFLICT 991 L -> F (IN REF. 4).
FT CONFLICT 1116 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 45.4%; Score 3088; DB 1; Length 1210;
Best Local Similarity 49.5%; Pred. No. 1.1e-157;
Matches 630; Conservative 168; Mismatches 363; Indels 112; Gaps 25;

QY 11 LLLALLPPGAA--STOVCTGTDMKLRUPASPTHLDMLRLHYQGVVQGNLELYLPTN 68
DB 14 LLTALCAAGALEKKVCCQGSNRLTQLGTGFEDHFLSLQRYNNCEVVLGNLEITYVORN 73
QY 69 ASLSFLQDIOEVQVYLIAHNVQVPLQRLIRVGTQLFEDNYALVLDNGDPLNNTTP 128
DB 74 YDLSFLTKIOEVAGVYLIALNTVERIPLENIQIIRGNALYENTYALAILSN----- 124
QY 129 VTGASPGRLRELQRLSLEILKGVLIQORNPLCQVDTILWKDI-----FHKQNLALTLI 184
DB 125 -YGTNRGLRELPRNLQELILIGAVRFNSNPILCNMDTIQWRDIVQNVFMSNMGMDL--- 180
QY 185 DTRNSRACHPCSPKCKSRGWGESSEDCOSLTRVTCAGGCA-RCKGPLPTDCHEQCAAG 243
DB 181 -QSHPPSCPKDPCPNCSGNGCWGGEENCQKTKIICAQCSHRRCRGRSPDCCHNQCAAG 239
QY 244 CTGPKHSDCLACLFHNSGICELCHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPVNY 303
DB 240 CTGPRESDCLVCKQFQDEATCKDTPPLMLNPTTYQMDVNPCKYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTLVCPHNOEVAEDTQRCCKSKPCARVCVGLGMQYIKANSKFTGITELE 363
DB 300 VVTDHGSCVRACGPDYIEV-EEDGIRCKCKDGPCKVCGIGIGIEFK-DTLSINATNIK 357
QY 364 -FAGCKKIFGSLAPLPESFDGDPASNTAPLOPELOQVFTELEITGVLYISAWPDSLPLD 422
DB 358 HFKYCTAISGDHLPLVAFKGDSTRTPTDPLELEILKVKETITGELLQAWPDNWTDL 417
QY 423 SVFQNLQVIRGRILHNGAYSLLTQGLGISWGLRSLRSLRELASGLALIHNTHLCPVHTVPW 482
DB 418 HAFENLEIRGRTRKQHGOFSLAVVGLNITSLGLRSLKEISDGDVLIISGNRLCYANTINW 477
QY 483 DQLFRQVIKANSKFIGITELECEGEGELACHQLCARGHCWGPPTQCVCNCSQFLRGQSCVE 542
DB 478 KCLFGTNPQTKIMNNRAEDCKAVNHVCNPLCSSEGCGWPEPRDCVSCQNVSRGECVE 537
QY 543 ECRVLQGLPREYVNAHCLPCHPCQPNQSVTCFGEADQCVACAHYKDPKPPFCVARCPS 602
DB 538 KCNILEGEPRFVENSECICHPECLPQAMNITCTGRGPNOCIOCAHYIDGPHCVKTCPA 597
QY 603 GVXPDLISYMPIWFPDEBAGCQPCPINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVGIL 662
DB 598 GINGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQGCVEWPSGPKPSIATGIUGGL 656
QY 663 LVVVGLGVVFGI-LIKRQOQKIRKYMRLLOETELVEPLTPSGAMPNQAOIRILKETELR 721
DB 657 LFIW-VALGILGFMRRRHVIRKTRLLQLERLEVEPLTPSGEAPNQAHRLILKETEFK 715
QY 722 KVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVAVGVSPI 781

```

```

DB 716 KIKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVAVGVSPI 775
QY 782 VSRLLGICLSTVOLVTQLMPEYGLLDHVRNRRGLSGSODLLNWCQIAKAGMSYLEDVRL 841
DB 776 VCRLLGICLSTVOLITQLMPEYGLLDYVRHKONIGSQYLLNWCQIAKAGMSYLEDRL 835
QY 842 VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRRPTH 901
DB 836 VHRDLAARNVLKTPQHVKITDFGLAKLGAEEYHAEKGVPIKMMMALESILHRIYTH 895
QY 902 QSDVMSGVTVTWELMTGAKPYDGPAREIPDLLEKGERLPOPPCTIDVYMWKCMWI 961
DB 896 QSDVMSGVTVTWELMTGSKPYDGPASDISILEKGERLPOPPCTIDVYMWKCMWI 955
QY 962 DSECRPRFRELVSFSEMRARDPQRFVVIQ-NEDLGPASPLDSTFYRSLEDDMDGLDVA 1020
DB 956 DADSRPKFERELLEFSKWARDPQRYLVIOQDERMHLPSPTDSNFYRALMDEDMEDVDA 1015
QY 1021 EYLVPOQGFPCPDPAAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAEPSPRIAPSEG 1080
DB 1016 DEYLIPOQGF-----NSPST-----SRTPLLSLS 1041
QY 1081 AGSDVFDGDLGMAAGKGLQSLPTHDPSPLOQRYSEDPTVPLPSET--DCYVAPLTCSPQPE 1138
DB 1042 ATSN-----NSTVACINRNGSCRVKEDAPLQRYSSDPTGAVTEDNIDDAFL-----PVPE 1091
QY 1139 YVNOPDVRPSPSPREGPLPAARPAAGATLERAKTSLPGKNGVWQDVAFAGGAVENPEYL- 1197
DB 1092 YVNO-SVPKPSAGSVQNPVHNOPLHP-----AFGRDLHYQN--PHSNAVGNPEYLN 1140
QY 1198 TPQGGAAAPHPHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPPSTFKGTP 1241
DB 1141 TAQ-----PTCLSGGNSPALWIKGSHQMSLDNPDYQDFFPKETKENGIFKG-P 1190
QY 1242 TAENPEYGLDVP 1254
DB 1191 TAENAEYLRVAPP 1203

RESULT 6
ERB4_HUMAN
ID ERB4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).
GN ERBB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RA "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
RA "A novel juxtamembrane domain isoform of HER4/Erbb4. Isoform-specific
RT tissue distribution and differential processing in response to
RT phorbol ester.";
RL J. Biol. Chem. 272:26761-26768(1997).

```





Db 531 ICIESCNLYDGEFREFENGSIQVCEDPQCEKMEGDLTCHGPGDNCTKCSHFQDGNVCV 590  
Qy 598 ARCSGVKPDLSYMPIKFPDEGACQPCINCHSCVDLDDKC-----PAEQ 646  
Db 591 EKCPDGLQGANF--IFKYADPDRECHPCNCTGCGNGPTSHDCIYYVPTWGHTSLPQHA 648  
Qy 647 RASPLTSISAVV--GILLVVLGVVGLILIKRQOKIRKYTWRLRLQSTELVEPLTSGA 705  
Db 649 R-TPL--IAAGVIGGLFVLVIGUTFAVYVRKSIK-KKRALRRFL-ETELVEPLTSGT 703  
Qy 706 MPNOAQMRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765  
Db 704 APNOAQLRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 763  
Qy 766 EILDEAVYMGVSGPYRLLGICLTSTVQLVLTOLMPYVGCILLDHVRENRLGSDLLNW 825  
Db 764 EFMDEALIMASMDPHLVRLLGLVCLSPITQLVTLMPHGCILLYVHEHKDNIGSQLLNW 823  
Qy 826 CMOJAKGMYLEDVRLVHRDLAARNVLKSNHVKITDGLARLLDDETEYHADGGKVP 885  
Db 824 CVQJAKGMYLEERLVRDLAARNVLKSNHVKITDGLARLLDDETEYHADGGKVP 883  
Qy 886 IKWMALESILRRRTHOSDVMYSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLP 945  
Db 884 IKWMALECIHYRKETHOSDVMYSYGVTVWELMTFGKPYDGIPTREIPDLLEKGERLP 943  
Qy 946 ICTIDVYMWKMWIDSECRPRELVSFERSMARDPQRFVITQNEP-LGPASPLDSTF 1004  
Db 944 ICTIDVYMWKMWIDADSRPKELAAEFPSMARDPQRYLVITQDGRMKLPSPNDSKF 1003  
Qy 1005 YRSLEDDDDGDLVDAEYLVPOQFFCPDPAPGAGGVHHRSSSTRSGGDLTGLLE 1064  
Db 1004 FQNLLEDDEDDMMDAEYLVLP-QAFNIPPP-----IYTSRARIIDSNRS-----EIGH 1051  
Qy 1065 PSEBEAPRS-----PLAP-SEGAGSDVDFDGLGMGA 1095  
Db 1052 PPPAYTPMSGNQFYVRDGGFAAEQGVVPYRAPYTPSTIPEAPVAGATAEIPDDSCCNGTL 1111  
Qy 1096 KGLQSLTHDPSLQRYSEDTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRPQ 1148  
Db 1112 RKPVAPHVQEDSSQRYSDPTVPAPERSPRGELDEGYTPMDRKPQYELNVE--- 1167  
Qy 1149 PPSPREGLPAARAGATLERAKTLSPCKNGVVDVAFGAVENPEYLTPOGGAAPQPH 1208  
Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPYHNASNG----- 1194  
Qy 1209 PPPA-----FSPAFNLYWDDQDPPERGA-- 1232  
Db 1195 PKAEDEVNPEPLNTFANTLGAEYLNKILNSMPEKAKAFDNPYWNHSLUPRSTLQ 1254  
Qy 1233 PPSTFKGTPT-----AENPEYL 1249  
Db 1255 HPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

RESULT 7  
ERB4\_RAT  
ID ERB4\_RAT STANDARD; PRT; 1308 AA.  
AC Q62956; Q922N7;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).  
GN ERB4 OR TYRO-2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=98221155; PubMed=9553078;  
RA Zhao Y.-Y., Sawyer D.R., Balliga R.R., Opel D.J., Han X.,

RA Marchionni M.A., Kelly R.A.;  
RT "Neuregulins promote survival and growth of cardiac myocytes.  
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult  
RT ventricular myocytes.";  
RL J. Biol. Chem. 273:10261-10269(1998).  
RN [2]  
RP SEQUENCE OF 848-901 FROM N.A.  
RC TISSUE=Sciatic nerve;  
RX MEDLINE=91222560; PubMed=2025425;  
RA Lai C., Lemke G.;  
RT "An extended family of protein-tyrosine kinase genes differentially  
RT expressed in the vertebrate nervous system.";  
RN Neuron 6:691-704(1991).  
RN [3]  
RP SEQUENCE OF 1031-1198 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuregulins and their putative receptors, ErbB2 and  
RT ErbB3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659(1997).  
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
CC RECEPTORS (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING  
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS  
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE  
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND  
CC HEART.  
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; AF041838; AAD08899.1; -;  
DR EMBL; U52531; AAC53051.1; -;  
DR HSP; P11362; IFGK.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR0109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 4.  
DR SMART; SM00219; TyrcK; 1.  
DR PROSITE; PS0107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00011; PROTEIN KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 25 POTENTIAL.  
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 652 675 POTENTIAL.  
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).





```
Db 590 IPRCPHGILGDDTL-IWKYADKMGQCQCHQNCQTQCGSGPLSGCRGD-IVSHSSILAVG 647
Qy 657 AVGILLVVLGVVFGILIKRRQKIRKYTHRRILQETELVEPLTPSGMNPQOMRLK 716
Db 648 LVSGLLITVIVALLIVLLRRRIK-RKRTICRLQEKELVEPLTPSGAENQAFRLK 706
Qy 717 ETELKRVKLGSGAGFYVYGIWIPDGENVPIPAIKVLRENTSPKANKETLDEAYVAG 776
Db 707 ETEFKDORVLGSGAGFYVYKGLWNPGENIRIPAIKVLRENTSPKANKETLDEAYVAG 766
Qy 777 VGSYVSRLLGICLTSTVQLVQMLPYGCLLDHVRNRLGSGQDLNMCQIAKMSYL 836
Db 767 VDHPVCRLLGICLTSAVLQVLTQMLPYGCLLDYVRQHQERICGQWLLNMCVQIAKMSYL 826
Qy 837 EDVPLVHRDLAARNVLKSNHVKITDGLARLIDIDETEHADGGKVPKIKMALESILR 896
Db 827 EERHLVHRDLAARNVLLKNPNHVKITDGLSKLLTADEKEYQADGGKVPKIKMALESILQ 886
Qy 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDYMIMV 956
Db 887 WTYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAKIASVLENGERLPOPPICITIEVYMIIL 946
Qy 957 KCMWIDSECRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMGD 1016
Db 947 KCMWIDPSSRRPRFRELVSFMRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMGD 1001
Qy 1017 LVDAEYVLVPOQGFCDPAPGAGGWHHRHSSTRSGGDLTLGLPSEEAAPRPLA 1076
Db 1002 VVDAEYLLPYKRI-----NRQGS-----EPC 1024
Qy 1077 PSEGAGSDVDFDGLGMAAGLQSLPHTDPSPLQRYSEDPTV-PLPSETDGYVAPLTCSP 1135
Db 1025 PPTGH-----PVRENSITLRNISDPTQNALEKLDGH----- 1056
Qy 1136 QPEYVQPDVRRQP-----PSPRE-----GPLP-AARPAGATLERAKTUSPKNGVVK 1182
Db 1057 --EYVNPQGSTSRSLSDIYNPNYEDLTGNGVPSLSQEAETNFSRPEYLTNQNSL-- 1112
Qy 1183 DVFAFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPRGPAPPSFTKGTPT 1242
Db 1113 -PLVSSGSMDDPDY---QAG-----YQAAF-----LPQTGAULTGNGMFLPA 1149
Qy 1243 AENPEYLG 1250
Db 1150 AENLEYLG 1157

RESULT 9
ERB3 HUMAN
ID -ERB3 HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR TransMembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
DR Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
```

FT	DISULFID	210	218	BY SIMILARITY.
FT	DISULFID	214	226	BY SIMILARITY.
FT	DISULFID	227	235	BY SIMILARITY.
FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	552	565	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	141	183	EILSGGVYIEKDKLCHMDTIDWRDIAEIVVNDGR SC -> GQFNPVSGLTQPQADWYLLDDDDPRLLTLLSASSK VPVTLAAV (IN SHORT ISOFORM).
FT	VARSPLIC	184	1342	MISSING (IN SHORT ISOFORM).
FT	CONFLICT	560	560	E -> G (IN REF. 2).
FT	CONFLICT	1064	1064	E -> G (IN REF. 2).
FT	SEQUENCE	1342	AA; 148097 MW; 7201ETP66CA374BD CRC64;	
Qy	Query Match	35.2%;	Score 2393.5; DB 1; Length 1342;	
Qy	Best Local Similarity	40.0%;	Pred. No. 1.5e-120;	
Qy	Matches 530; Conservative	194;	Mismatches 445; Indels 155; Gaps 35;	
Qy	10	GLLALLUPGAA--STQVCTGDMKRLRPLASPETHLMDLRLHYOGCVQVQGNLELTLYPT	67	
Db	11	GLLPSLARGSEVGNQAVCPGLNGLSVTGDENQYQTLKLYERCEVMGNLEIVLTGH	70	
Qy	68	NASLSFLQDIQEVGYVLIAHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTT	127	
Db	71	NADLSFLOWIREVTGYVLVAMNEFTLPLNLRVVRGTQYVDGKFAIFVM-----LNYNT	125	
Qy	128	PVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILWKDIFHKNNQALTLIDTN	187	
Db	126	-----NSSHALRQLRLTQTEILSGGVYIEKDKLCHMDTIDWRDIAEIVVND	178	
Qy	188	RSRACHPCSPMGKSGRCWGSSEDCQSILTRTVCAAGC-ARCKGPLPTDCHEQCAAGCTG	246	
Db	179	NGRSCPPCHEVCKG-RCWGPGEDECQTLTKTICAPQCNGHCFCFNPQCCHDECAGCGSG	237	
Qy	247	PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLYST	306	
Db	238	PQDTCDFACRHFNDSGACVRCRQPLVYNKLTFOLEPNPHTKYQYGVGVASCPHNFV-V	296	
Qy	307	DVGSCTLVCPHNOEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKF--IGITELE-	363	
Db	297	DQTSCTVRACPPDKMEVD-KNGLKXCEPCGGLCPKACEGTG-----SGSRFTQVDSNIDG	350	
Qy	364	FAGCKKIFGSLAFIPESFDGDPASNTAPLOPEQLOQVFTLEBEITGYLYISAWPDSLPDLS	423	
Db	351	FVNCTKLGNLDFLITGLNGDPMHKIPALDPKLNVRFTVREITGYLYIQSWPPHMFNS	410	
Qy	424	VFNQVQIRGRILHNGAYS-LTLOGLGISWGLRSLRELGLSGLALIHNTLHCFVHTVPV	482	
Db	411	VFSNLTITGRSLYNGRCFSLIKMNLNVTSLGFRSLKEISAGRIYISANKQLCYHHSNLA	470	

Qy	483	DQLFRQYIKANSKFIGITE-----LECVGEGELACHQLCARGHCWGPQTQVCNC	531
Db	471	TKVLR-----GPTTEERLDIKHNRPRRDCVAEGKVCPLCSCSSGCGWGPQCLSC	520
Qy	532	SQFLRGQECVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHYK	591
Db	521	RNYSRGVGVTHCNFLNGEPREFAEACFSCHPCECPMEGTATCNGSGSDTCAQCAHFR	580
Qy	592	DPFPCVARCPSGVKPDLSYMPIWKFPDBEGACQPCPINCTHSCVDLDDKGCQPAEQA	648
Db	581	DGPHCVSSCPHGVLG--AKGPIYKYPDVONECRCHENCTQCGKGPGLQDCLGQTLVLIG	638
Qy	649	-SPLTSIVSAVVGILLVVLGVFGIILTKRQOKIR-KYTMRLRLQETELVEPLTPSGAM	706
Db	639	KTHLTMAITVIAG--LVVIFMMLGGTFLYWRGRRIQNKRAMRYLERGESIEPLDPS-EK	695
Qy	707	PNQAQMRILKETELRKVKVLSGAGFTYKGIWIPDGENVKIPVAIKULRENTSPKANKE	766
Db	696	ANKVLARIFKTELURKLVKLSGVGFTVHKGVWIPGESIKIPVCIKVIEDKSGRQFOA	755
Qy	767	ILDEAYVMAGVGSYPVSRLLGICLTSTVQLVTQLMPPYGLCLLDHVRENRRGLSGDLLNWC	826
Db	756	VTDEMLAIGSLDHAHIVRLGLCGSSQLQVTLPLGSLLDHVQRHGAQGPQLLWNG	815
Qy	827	MQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP	886
Db	816	VQIAKGMVYLEEHGMVHRNLAAARNVLLKSPSQVQVADGVADLPPDDKQLLYSEAKTPI	875
Qy	887	KWMALESILARRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPI	946
Db	876	KWMALESIFHGKYTHQSDVMSYGVTVWELMTFGAEPYAGLRLAEVPLLEKGERLAQPI	935
Qy	947	CTIDVYMWKCMWIDSECRPPRELVESESRMARDPQFVVIQNEIDLGA---SPLDST	1003
Db	936	CTIDVYMWKCMWIDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHG	994
Qy	1004	FYRSLLEDDMDGLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGL	1063
Db	995	LTKKLEVELEPELDDLDLEAED-----NLATTLGSALSPLV	1035
Qy	1064	EP-SEBAPRSPLAPSEGAGSDVFDGLGMAAGLOSPLTHD-PSPLQRYSESDFTVPLP	1121
Db	1036	GTLNRRPGSOLSLSPSSGY-MPMNQNLGSECSAVSGSSERCPRPVSLH-----BMP	1088
Qy	1122	-----SETGYVA-----PLTCSPOPE-----YUNQPDVPRQPPSPREG	1155
Db	1089	RGCLASSESSEGHVTGSEAELOEKVSMCRSRSRSPRGRGDSAYHSQRHSLLTPVPLSP	1148
Qy	1156	P-----LPAARPAGATLERAKTLSP-GKNGVV-----KDVPAFGAVENPEYL	1197
Db	1149	PGLEEDVNGVMPDTHLKGTPSSREGTSSVGLSSVLGTDEED-----EYEYM	1200
Qy	1198	TPQGGAAPQHPHPPAFSPAFDNLVYWD-----QDPPERGAPSTFKGPTTAEN	1245
Db	1201	NRRRHSP-PHPRPPSSLEELGYEYMDVGDLSASLSTQSCPLHPVIMPTAGTTPDED	1259
Qy	1246	PEYL 1249	
Db	1260	YEYM 1263	

RESULT 10

ERB3 RAT

ID -ERB3 RAT STANDARD; PRT; 1339 AA.

AC Q62799; Q62955;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)

DE (c-erbB3).

GN ERBB3.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;





```

Db      520 CRNYSREGVCTHNFQGEPRFVHEAQCFSCHEPLMEGTSYTCNGSGSDACARCAHF 579
Qy      591 KDPFFCVARCPGSKVPLDLSYMPKWKFPDEGACOPCPINCTHSC--VDLDDKGPAPQRA 648
Db      580 RDGPHCVNSCPHGILG--AKGPIYKPDQAQNECPCHENCTQCGNFELQDCLGQAEVLM 637
Qy      649 SPLTSIVSVAVGILLVVLGVVFGILKROQKIR-KYWRRLLOETLEVEPLTPSGAMP 707
Db      638 SKPHLVIAVTVG--LAVILMLGSGFLYWRGRRITQNKRAMRRYLERGESIEPLDPS-EKA 694
Qy      708 NOAQMRILKTELKRVKVLGSGAFGTGVYKGIWIPDGENVKIPVAIKVLENTSPKANKI 767
Db      695 NKVLARIFKTELKRVKVLGSGVFGTVHKGWIPEGESIKIPVCIKVIDKSGRQSQAV 754
Qy      768 LDEAYNVAGVSPVSRLLGICLTSTVQLVTPQYGLLDHVRNRRGLSGQDLNWM 827
Db      755 TDHMLAVGSLDHAHIVRLGLCPGSSQLVTOYPLGLSLLDHVKQHRETILGPQLLNWV 814
Qy      828 QIAKGSYLEDLVRLVHRDLAARNVLKSPHVKITDFGLARLLDIDETEHADGKGVPIK 887
Db      815 QIAKGMYLEEHSVHRDLARNVMLKSPSQVQVADFGVADLLPPDDKQLHSEAKTPIK 874
Qy      888 WMALESLRRRFTHQSDVWSVGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPIC 947
Db      875 WMALESHFGKYTHQSDVWSVGVTVWELMTFGAEPYAGLRLAETPDLLEKGERLAQPOIC 934
Qy      948 TIDYIMVIMKWMIDSCRFRELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYS 1007
Db      935 TIDYVMVIMKWMIDENIRPTFELANEFTRMARDPPRYLVIKRAS-GPGTP--PAEAPS 991
Qy      1008 LLEDDMGDLVDAEYLVPOGFPFCDPAPGAGGVHRRSSSTRSGGGDLTLGLPSE 1067
Db      992 VLTTKEL-----QEALELEP-----AGSDVFDGDLGCAAGLQ 1017
Qy      1068 E-----EAPRSLAPSEG-----AGSDVFDGDLGCAAGLQ 1099
Db      1018 EGLATSLGSLSLPTGLTLPGRSQSLSSSGVMPNQSLGACLDLSAVLGREGFSR 1077
Qy      1100 SLPHDPSLPQRYSEDTPVLPSETDGVY----APL-----TC-----SPOPE----Y 1139
Db      1078 PISLH-PIPRGR-----PASESEGHVTGSEAEQLQKVSVCFSRSRSPRGRDSAY 1129
Qy      1140 VNPQDVRPQPPSPREGP-----LPAARPAGATLERAKTLSP-GKNGVY----- 1181
Db      1130 HSQRHSLTLTPTPLSPGLEEDGNGYVMPDTHLRGASSRREGTLSSVGLSLVLTREED 1189
Qy      1182 KDVFAGCGAVENPEYLTPOGGAAPQPHPP 1210
Db      1190 ED-----EEYEMNRKRGSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P04412; O61501; Q9W2G0; P91868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
that several genetically defined classes of alleles cluster in

```

```

RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
[2]
RN REVISIONS.
RP Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
[4]
RN SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RP STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
[5]
RN SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
ANALYSIS.
RP MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
[6]
RN SEQUENCE FROM N.A. (ISOFORM TYPE I).
RP STRAIN=Berkely;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dinkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Waskarman D.A., Weinstein G.M., Weissenbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,
Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[7]
RN SEQUENCE OF 959-1078 FROM N.A.
RP

```





```
Db 543 RPAIQKEPQKVVNNENLR-----ADL-CEKNGTICSDQCNEGCMAGTDQCLCTCNFN 597
Qy 536 RGCEVCEECVLOGLPREYVNAHCLPCHPECPONGSVTCFCGPEADQCACAHYKDPFP 595
Db 598 FNGTCTADCGYISNAYK--FDNRCKICHPECR-----TCNGAGADHCBCEVHRDQGH 649
Qy 596 CVARCP-----SGVK----- 605
Db 650 CVSECPKNKYNDRGVCHRECHATCDGCTGPKDTIGIGACTTCNLAIINNDATVKRCLLKD 709
Qy 606 --POLSNMPTWKPF--PDEEGACQP-----CPI-----NCTH----- 632
Db 710 KCFD-GY--FWEYVHPOEQSLKPLAGRAVCRKCHPLCELCNTNYGHEQVCSKCTHYKRR 766
Qy 633 -----SC-----VDLDDKG----- 641
Db 767 EQCETECPADHYTDEORECFQHPCEGCTGPGADDCCKSRNFKLFANETGPNVNTM 826
Qy 642 -----CPAEOR-----ASPLTS-----IVSAVVGILLVVVLG 668
Db 827 FNCTSKCPLEMRHVNYQYTAIGPYCAASPPRSSKITANLDVNMFIITGAVLVPTCILC 886
Qy 669 VFGILLIKRQOKIRKVT--MRLLQETELVEPLTPSGAMPNQAMQMRILKETELRKVKVL 726
Db 887 W--TYICRQKQKAKETVMWTWALSGCEDSEPLRPSNIGANCLKRIVKDAELRKGGLV 944
Qy 727 GSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKELDEAYVMAGVSPYSRL 786
Db 945 GMAFGVYKGVVPEGENVKIPVAIKELKSTGAESSEFLEAYIMASEEHNVLKLL 1004
Qy 787 GICLTSTVOLTPMPYGCILLDHVRENRLGSDLLNMCQIAGMSYLEDLVRLVHRDL 846
Db 1005 AVCMSQMLITQLMPLGCLLDVYNNRDKIGSKALLNWSQIAGMSYLEKELVHRDL 1064
Qy 847 AARNLVKSPNHVKITDFGLARLLDDEYHADGKGVPIKWALESILRRRFTHQSDVW 906
Db 1065 AARNVLVQTPSLVKITDFGLAKLLSSDSNEYKAAGKMPKWLALCEIRNRVFTSKDVW 1124
Qy 907 SYGVTVWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVVMVIMVKCMIDSECR 966
Db 1125 AFGVTIWEITTFQRPHENIPAKDIPDLIEVGLKLEQPEICSDIYCTLLSCWHLDAHR 1184
Qy 967 PRFRELVSFMRARDPQRFVITONEDLG--PASPLDSTFYRSLLED--DMGDLVDAE 1021
Db 1185 PTEKQLTTPFAEPADPGRYLAIPGDKFTRLPA-----YTSQDEKDLIRKLAPTTDGS 1237
Qy 1022 EYLVPOOGFCPPAPCAGGVVHRHRSSTRSGGDLTLGLPSEBEAP-----RSPL 1075
Db 1238 EAIKAPDDYLOPKAAPGPS-----HRTDCT-----DEMPKLNRYCKDPS 1276
Qy 1076 APSEGAGSDVFDG---DLGMAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETGYVAPLT 1132
Db 1277 NKNSSTGDDERDSAREVGVGNLR-----LDLPVDEDDYLMPT-T 1314
Qy 1133 CSPQEVYVNOPDVRPOPPSPREGPLPAARPAATLERAKTLPSPKNGVWVDVFAFGAVE 1192
Db 1315 CQPGPNNNNMN-----NPNQNNMAVGAAGVM-----DLIGVPVSVD 1353
Qy 1193 NPEYL-----TPQGAAPQPH-----PPAFSP-AFDNLYYWD 1224
Db 1354 NPEYLNQAOTLGVGESPIPTQTIGIPVMGGPGTMEVKVPMPPGSEPTSSDHEYND 1408
```

## RESULT 12

```
ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
```

```
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
processing and promoter insertion result in expression of an
amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -|- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -|- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
PROTEIN.
CC -|- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
EMBL: M10066; AAA48763.1; ALT_INIT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVFVLV.
DR HSP: P11362; IFGK.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00219; TyrcK; 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00109; PROTEIN KINASE TVR; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR Transfaser: Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 1.6e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 587 CAHYKDPPEFCVAPCPGVRKPDLSYMPIWKPDEGACQPCINCTHSCVDLDDKGCFAEQ 646
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VWKYADANAVCOLCHPNCTCGKGPGLGCP--- 58
Qy 647 RASPLTSIVSAVV-GILLVVVLGVVFGILKRRQOKIRKYTMRLLOETELVEPLTPSGA 705
Db 59 NGSKTFSIAAGVVGGLLCLVVVGIGLGLYLR--HIIVKRTLLRLLQERLVEPLTPSGE 117
Qy 706 MPNQAMRILKETELRKVKVGLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765
Db 118 APNQAHRLILKETEPFKVKVGLSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 766 EILDEAYVMAGVSPYSRLGLICLTSTVOLTPMPYGCILLDHVRENRLGSDLLN 825
Db 178 EILDEAYVMASVDNPHVCRLLGLICLTSTVOLTPMPYGCILLDHVRENRLGSDLLN 237
Qy 826 CMQIAGMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDEYHADGKVP 885
```

```
Db 238 CVOIAKGMNLEERRLRVHRLDAARNVLVKTQHVKVTDFGLAKLLGADEKEYHAEGGKVP 297
Qy 886 IKMALESILRRRFTHQSDVMSGYVTVMELMTFCGAKPYDGGIPAREIPDLLEKGERLP 945
Db 298 IKMALESILHRIYTHQSDVMSGYVTVMELMTFCGAKPYDGGIPAREIPDLLEKGERLP 357
Qy 946 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQREVVQI-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQREVVQI-NEDLGPASPLDSTF 417
Qy 1005 YRSLEDDDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRSP-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCID-----RNGQHPVREDSFVQRYSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQVRPQPPSPREGPLPAARPAGATLERAKTLP 1177
Db 496 FLEESIDDDGFL-----PAPEYVQ--LMPKKPS-----TAMVQ 526
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQHPHPPAFSPADNLYY 1222
Db 527 NQIYNNISLTAISKLPMDSRQYONSHSTAVDNPYL-----NTNOSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RGAPSTFTGTPTAENPEYLGLDVP 1254
Db 579 WIOGNGHQLNDNDYQODFLPNETKPNGLLKVPAENPEYLVRAAP 625

RESULT 13
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=63113229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RL gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debierre B., Henry C., Benaisse M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RL new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBCDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 4.7e-84;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKDPFVCVARGSPGVKPDLSYMPIWKFDEGACQPCPINCSTHSCVDLDDKGCPAEQ 646
Db 3 CAHFIDGPHCVKACPAVGLGENDTL-VRYKADANAVCOLCHPCNCTRGCKGLEGCP--- 58

Qy 647 RASPLTSIVSAV-GILLVVVGVVFGILIKRQOKIRKYTMRLRLQSTELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVGGLLCLVVGIGLYLRRR-HIVKRTLRLRLQSTELVEPLTPSGE 117

Qy 706 MPNOAQWRIKTELKRVKVLGSGAFGVYKGIWIDGENVKIPIVAIKVLENTSPKANK 765
Db 118 APNOAHRIILKETEFKKVKVGLSGAFGTIYKGLWIPGEKVKIPIVAIKELRETSPKANK 177

Qy 766 EILDEAYVMAGVSGPYVSRILGICLTSTVQLVTQMLPYGCLLDHVRENRGLSGDGLNW 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDVIREHKDNGISYLLNW 237

Qy 826 CMQIAKMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAKGMNLEERRLRVHRLDAARNVLVKTQHVKVTDFGLAKLLGADEKEYHAEGGKVP 297

Qy 886 IKMALESILRRRFTHQSDVMSGYVTVMELMTFCGAKPYDGGIPAREIPDLLEKGERLP 945
Db 298 IKMALESILHRIYTHQSDVMSGYVTVMELMTFCGAKPYDGGIPAREIPDLLEKGERLP 357

Qy 946 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQREVVQI-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMWKMWIDSECRPRELSEFSESRMARDPQREVVQI-NEDLGPASPLDSTF 417

Qy 1005 YRSLEDDDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHSSSTRSGGDLTLGLE 1064
Db 418 YRTLMEEEDMEDIVDAEYLVPHQGF-----NSPST----- 449

Qy 1065 PSEEEAPRSP-----APSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVP 1119
Db 450 -----SRTPLLSLSLTSNNSATNCID-----RNGQHPVREDSFVQRYSDPTGN 495

Qy 1120 LPSET--DGYVAPLTCSPQPEYVNPQVRPQPPSPREGPLPAARPAGATLERAKTLP 1177
Db 496 FLEESIDDDGFL-----PAPEYVQ--LMPKKPS----- 524

Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQHPHPPAFSPADNLYY 1222
Db 527 NQIYNNISLTAISKLPMDSRQYONSHSTAVDNPYL-----NTNOSPLAKTVFESSPY 1218
```

```

Db 525 --VQNIYFISLTAISKLPWD3RQNSHSTAVDNPEYL-----NTQSPKLTVP 574
QY 1219 NLYWQDPPPERGAPPSTFGTPTAENPEY 1248
Db 575 SSPYIQSGNHQ-----INLNDPDY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OS Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87064458; PubMed=2878364;
RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells."
RT Mol. Cell. Biol. 6:1751-1759 (1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M13179; AAA2401.1; -
DR PIR; A25231; TVFVEB.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00669; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TVR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B65D CRC64;

Query Match 24.0%; Score 1630; DB 1; Length 540;
Best Local Similarity 55.1%; Pred. No. 3 3e-80;
Matches 341; Conservative 69; Mismatches 119; Indels 90; Gaps 14;

QY 587 CAHYKDPFVCARCSGVKPDLSYMPFIWPPDEGACPCPINCTHSCVDLDDKGCQPARQ 646
Db 3 CAHFIDGPHCVKACPAVLGENDTL-VKMYADANAVCQLCHPNCRTGCKGPGLEGCP--- 58
QY 647 RASPLTSIVSAVY-GILLVVLGVFGILIKRQKIRKYTWRRLLQETELVEPLTPSGA 705
Db 59 NGSKTPTSAAGVVGGLLVVGLGIGLYLRRR-HIVRKRTLRLQLRELVEPLTPSGE 117
QY 706 MPNQAMRILKELTELKRVKVLGSGAFGTGKGIWIPDGENVKIPVAIKVLRNTSPKANK 765
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

```

```

Db 118 APQAHURLIKETEFKKVKVLGFGAFCTVYKGLWIPGEKVTIPVAIKELREATSPKANK 177
QY 766 EILDEAYVMAGVGSPPYVSRLLIGICLTSTVOLVLTQMPYGCGLLDHVRNRRGLSGODLLNW 825
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 178 EILDEAYVMASVDNPHVCRLLIGICLTSTVOLITQMPYGCGLLDYIREHKONIGSQYLLNW 237
QY 826 CMQIAKMSYLEDLVRLVHROLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 238 CVQIAKGMNYLEERHVMVHRLAARNVLKTPQHVKITDFGLAKQLGADKEYHAEKGGKVP 297
QY 886 IKWMALESILRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPP 945
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 298 IKWMALESILHRYITHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEKGERLPOPP 357
QY 946 ICTIDVYIMVKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQ-NEDLGPASPDLSTF 1004
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 358 ICTIDVYIMVKCMWSADSRPKRELIAEFSKWARDPPRYLVITQGERMHLPSPTDSKF 417
QY 1005 YRSLLDDDDGMDLVDAEEYLVPOGGFCPPDPAPGAGGMVHRHRSSTRSGGDLTLGLE 1064
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 418 YRTLMEEDMEDIVDAEVLVPHQGF-----NSPST----- 449
QY 1065 PSEEEAPRSL-----APSEGAGSDVFDGDLGMAAKGLQSLPHTDPSLQRYSEDTVP 1119
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 450 -----SRTPLSLSLTSATNSNATNCIDRNG-----H----- 476
QY 1120 LPSETDGYVAPLTCSPOPEYVQPDVRPQPPSPREGPLPAARPAGAT-LERAKTLPSPGN 1178
||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db 477 -PVREGDL-----PAPEYVQ--LMPKFPSTAMVQNIYVILSTAIKLPIDSRVON 527
QY 1179 GVVKDVFAGGAVENPEYL 1197
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
GN (Fragment).
DE EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha".
RL Mol. Cell. Biol. 8:1970-1978 (1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; M20386; AAA48760.1; -  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF01030; Recept\_L\_domain; 2.  
DR SMART; SM00261; FU; 4.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; PARTIAL.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; PARTIAL.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; PARTIAL.  
KW Transmembrane; Glycoprotein; Receptor; Signal; Transphosphorylation.  
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 30  
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.  
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 655 667 POTENTIAL.  
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 197 206 BY SIMILARITY.  
FT DISULFID 201 214 BY SIMILARITY.  
FT DISULFID 222 230 BY SIMILARITY.  
FT DISULFID 226 238 BY SIMILARITY.  
FT DISULFID 239 247 BY SIMILARITY.  
FT DISULFID 243 255 BY SIMILARITY.  
FT DISULFID 258 267 BY SIMILARITY.  
FT DISULFID 271 298 BY SIMILARITY.  
FT DISULFID 302 314 BY SIMILARITY.  
FT DISULFID 318 333 BY SIMILARITY.  
FT DISULFID 336 340 BY SIMILARITY.  
FT DISULFID 513 522 BY SIMILARITY.  
FT DISULFID 517 530 BY SIMILARITY.  
FT DISULFID 533 542 BY SIMILARITY.  
FT DISULFID 546 562 BY SIMILARITY.  
FT DISULFID 565 581 BY SIMILARITY.  
FT DISULFID 569 589 BY SIMILARITY.  
FT DISULFID 592 601 BY SIMILARITY.  
FT DISULFID 605 627 BY SIMILARITY.  
FT DISULFID 630 638 BY SIMILARITY.  
FT DISULFID 634 646 BY SIMILARITY.  
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT NON TER 703 703  
SQ SEQUENCE 703 AA; 77427 MW; AFFDE11B735A690 CRC64;

Query Match 23.1%; Score 1575; DB 1; Length 703;  
Best Local Similarity 44.2%; Pred. No. 3.8e-77;  
Matches 313; Conservative 111; Mismatches 254; Indels 30; Gaps 14;  
Qy 8 RWGLLLALLPPGAA-----STQVCTGDMKLRLPASPETHLDMLRLHYGCGVVGQNGLE 61  
Db 13 RGAIVLVLLGLGVALCSAVEKKVCGQTNNKLTQLGHVEDHFTSLQRMYNCEVVLNLE 72  
Qy 62 LTYLPTNASLFLQDIQEVQGVYLIAHNQVQLRIRIVRGTLQFEDNYALAVLNGD 121  
Db 73 ITVYEHNRDLTFLKTIQEVAGYVLIAMNVDVPLENLQIRGNVLYDNSFALVLSNH 132  
Qy 122 PLNNTTPTVGTGSPGLRELQRLSRTEILKGVLIQRNPOLCYQDTILWKDIFHKNQAL 181

Db 133 -MNKTQ-----GLRELPMKRLSEILNGGVKISNNPKLCNMDTVLWNDIITSRK-PL 182  
Qy 182 TLID-THRSRACHPCSPKRCWGSSESSEDCQSILTRVCAGGCA-RCKGPLPTOCCHQ 239  
Db 183 TVLDFASNLSSCPKCHPNCTEDHCWGAGBQNCQTLTKVICAQCCGRGKVPSCCHNQ 242  
Qy 240 CAAGCTGPKGSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 299  
Db 243 CAAGCTGPRESDCLACRFRDDATCKTCTPPLVLYNPTYQMDVNPPEKYSFGATCVREC 302  
Qy 300 PNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGLGMOYIKANSKFIGI 359  
Db 303 PHNVVVTDHGSCVRSCTNTDYTEV-EENGVRKCKCDGLCKVCNGIGIGELKGLS-INA 360  
Qy 360 TELE-FAGCKKIFGSLAPLSPESFGDPPASNTAPLOPEQLQVETLEETGVLVISAAPDS 418  
Db 361 TNIDSFKNCTKINGDVSILPVAFLDGAPTCTLPLDPKLDVFTVKEISGFLLIQAWPDN 420  
Qy 419 LPDLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGSGLALIHHTHLCFVH 478  
Db 421 ATDLIYAFENLEIIRGRTKQHQYSLAVNVNLIQSLGLSLKEISDGDIAIMKNVLCYAD 480  
Qy 479 TVPMDQLFRQYIKANSKFIGITELECVGEGEACHQLCARGHCWGPGTQCVCNSQFLRQ 538  
Db 481 TMNWSLFAFQSQKTKIIQNNKNDCTADRHVCDPLCSDVCGWCGPGPHCFSCRFESRQK 540  
Qy 539 ECVEECRVLQGLPREYVNAHCLCHPECQPNQ---SVTCFGEADOCVACAHYKDPFF 595  
Db 541 ECVKQCNILOGEPRERFERSKCLPCHSECLVQNSTAYNTTCSGPGDHCMAHFIDGPH 600  
Qy 596 CVARCPGSKPDLSPYMPIWPKFDEEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSTV 655  
Db 601 CVKACPAVLGENDTL-VWKYADANAVCOLCHPNCTRCCKGPGLEGCP---NGSKTPSIA 656  
Qy 656 SAVV-GILLVVVLGVWFGLIKRRQQKIRKYTMRLLOETELVEPLTP 702  
Db 657 AGVVGGLLCLVVGVLGIGLYLRRR-HIVRKRTLRLLQLRELVEPLTP 703

Search completed: July 22, 2003, 08:46:30  
Job time : 20.2304 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds  
(without alignments)  
4403.399 Million cell updates/sec

Title: SEQ4-369-383-12

Perfect score: 6815

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq\_101002.\*

```
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	6659	97.7	1255	21	Human heregulin 2
2	6659	97.7	1255	22	Human tyrosine kin
3	6659	97.7	1255	22	HER2 transgene pla
4	6659	97.7	1255	23	Human HER2 (ErbB2)
5	6653	97.6	1255	17	HER-2/neu protein.
6	6653	97.6	1255	20	Human HER-2/neu on
7	6653	97.6	1255	21	Amino acid sequenc
8	6653	97.6	1255	21	Human HER-2/neu pr
9	6653	97.6	1255	22	Amino acid sequenc
10	6653	97.6	1255	22	Human HER-2/neu pr

11	6653	97.6	1255	23	AAE24067	Human Her-2 protei
12	6653	97.6	1255	23	AAE20479	Human Her-2/neu pr
13	6653	97.6	1255	23	AAAM51143	Human Her-2/neu On
14	6653	97.6	1255	23	AAU77114	Human Her-2/neu po
15	6614	97.1	1433	14	AAAR39568	Sequence of c-erbB
16	6489	95.2	1223	23	AAAR98923	Human breast cance
17	6336	93.0	1200	21	AAAB21208	Human HER-2/neu pr
18	5869.5	86.1	1256	21	AAAB21199	Rat Her-2/neu prot
19	5869.5	86.1	1256	23	AAAM51144	Rat Her-2/neu onco
20	5846.5	85.8	1256	21	AAAB21206	Mouse Her-2/neu pr
21	5846.5	85.8	1256	22	AAAG2860	Amino acid sequenc
22	5846.5	85.8	1256	23	AAAM51151	Mouse Her-2/neu on
23	4739	69.5	919	21	AAAB21203	Human HER-2/neu fu
24	4739	69.5	919	23	AAAM51148	Her-2/neu extracel
25	3996.5	58.6	920	23	AAAM51152	Mouse Her-2/neu ex
26	3996.5	58.6	926	23	AAAM51153	Mouse Her-2/neu ex
27	3623	53.2	712	21	AAAB21204	Human HER-2/neu fu
28	3623	53.2	712	23	AAAM51149	Her-2/neu extracel
29	3477	51.0	782	18	AAAI9764	Her2-GM-CSF immuno
30	3475	51.0	653	21	AAAB21200	Extracellular HER-
31	3475	51.0	653	23	AAAM51145	Human Her-2/neu on
32	3437	50.4	645	22	AAAB60408	Human ErbB2 oncopr
33	3437	50.4	645	22	AAAB61593	Human ErbB2 extrac
34	3372	49.5	951	21	AAAY44993	DC8scfv-erbB2EC fu
35	3269	48.0	624	11	AAAR08222	Extracellular port
36	3097	45.4	1210	21	AAAB19259	Amino acid sequenc
37	3097	45.4	1210	21	AAAY50616	Human EGF receptor
38	3097	45.4	1210	23	AAE23019	Human Her-1 protei
39	3097	45.4	1210	23	AAAM50768	Human epidermal gr
40	3095	45.4	1210	22	AAAB68420	Amino acid sequenc
41	3084	45.3	583	23	AAE20483	Human protein for
42	3084	45.3	587	23	AAE20481	Human protein for
43	3083	45.2	589	23	AAE20484	Human protein for
44	3083	45.2	600	23	AAE20482	Human protein for
45	3056	44.8	1210	23	ABPS1768	Human epidermal gr

#### ALIGNMENTS

```
RESULT 1
RAY2620
ID AAY92620 standard; Protein; 1255 AA.
AC AAY92620;
XX
XX 10-AUG-2000 (first entry)
XX
XX Human heregulin 2 (Her2).
DE
XX Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Domain 1..173
FT /label= N-terminal
FT /note= "mature polypeptide"
FT Region 5..25
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 59..73
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 103..117
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Region 149..163
FT /label= insertion region
FT /note= "suitable for foreign epitope insertion"
FT Domain 174..323
```

FT /label= Cysteine\_rich\_domain  
 FT 210..224  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 250..264  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 324..483  
 FT /label= Ligand\_binding\_domain  
 FT 325..339  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 369..383  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 465..479  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 484..623  
 FT /label= Cysteine\_rich\_domain  
 FT 579..593  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 624..654  
 FT /label= Transmembrane\_domain  
 FT 632..652  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 653..667  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 655..1010  
 FT /label= Tyrosine\_kinase\_domain  
 FT 661..675  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 695..709  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 710..730  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 1011..1235  
 FT /label= C-terminal\_domain  
 FT  
 PN WO200020027-A2.  
 XX  
 PD 13-APR-2000.  
 XX  
 PF 05-OCT-1999; 99WO-DK00525.  
 XX  
 PR 05-OCT-1998; 98DK-0001261.  
 PR 20-OCT-1998; 98US-0105011.  
 XX  
 PA (MEBI-) M & E BIOTECH AS.  
 XX  
 PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
 PI Gautam A, Birk P, Karlsson G;  
 XX  
 DR WPI; 2000-349917/30.  
 DR N-:SDB; AAA09455.  
 XX  
 XX Inducing immune responses to weakly immunogenic, tumor associated  
 FT peptide antigens for the treatment of breast and prostate cancer  
 XX  
 PS Claim 62; Page 193-198; 220pp; English.  
 CC  
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
 CC response. Subdominant CTL epitopes, antibody binding regions and  
 CC cysteine residues involved in disulfide bonds are preserved in the  
 CC immunogenized forms. Regions suitable for the insertion of foreign T  
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
 CC cell-associated peptide antigens (PA) such as those associated with  
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
 CC The method comprises effecting simultaneous presentation by antigen  
 CC producing cells (APCs) of the animals immune system of: (1) at least 1  
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
 CC B-cell group derived from the cell-associated PA; and (2) at least 1  
 CC first T helper cell group which is foreign to the animal. Analogues of  
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
 CC part of all known and predicted CTL and B-cell epitopes of the respective  
 CC PA and including at least one foreign T helper epitope are also claimed.  
 CC The method is used to treat prostate, prostate/breast or breast cancer  
 CC when the PA is human PSM, FGF8b and Her2, respectively.  
 XX

Seq Sequence 1255 AA;

Query Match 97.7%; Score 6659; DB 21; Length 1255;  
 Best Local Similarity 97.8%; Pred. No. 0;  
 Matches 1227; Conservative 7; Mismatches 21; Indels 0; Gaps 0;

QY 1 MELAALCEWGLLLALLPFGAASVCTGCTDMKRLPASPETHLDMRLHYQCQVQGNL 60  
 DB 1 MELAALCEWGLLLALLPFGAASVCTGCTDMKRLPASPETHLDMRLHYQCQVQGNL 60  
 QY 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120  
 DB 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVRQVPLQRLIRVGTQLPEDNYALAVLDNG 120  
 QY 121 DPLNNTTPTVGASPGGLREQLRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180  
 DB 121 DPLNNTTPTVGASPGGLREQLRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKNNQLA 180  
 QY 181 LTLIDTNRSRACHPCSPMKCGSRWGESSEDCQSILTRTVACGGCARCGPLPTDCHEOC 240  
 DB 181 LTLIDTNRSRACHPCSPMKCGSRWGESSEDCQSILTRTVACGGCARCGPLPTDCHEOC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPLNQVETADGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360  
 DB 301 YNYLSTDVGSCTLVCPLNQVETADGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360  
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGLYIISAWPDSLP 420  
 DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGLYIISAWPDSLP 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGLSRLSELGSLALIHNTLHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTLOGLGISWLGLSRLSELGSLALIHNTLHLCFVHTV 480  
 QY 481 PWDQLFRNPQOALLHTANRPEDECVGEGLACHQLCARGHCGPGTQCVCNSQFIRGQBC 540  
 DB 481 PWDQLFRNPQOALLHTANRPEDECVGEGLACHQLCARGHCGPGTQCVCNSQFIRGQBC 540  
 QY 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600  
 DB 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600  
 QY 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPQASPLTISIVSAVVG 660  
 DB 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPQASPLTISIVSAVVG 660  
 QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTMERLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 QY 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
 DB 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780

QY 781 YVSRLLGICLTSTVOLVLTQMPYGCGLLDHVRNRRGLSGQDLNLMCMQIAKMSYLEYDR 840  
Db 781 YVSRLLGICLTSTVOLVLTQMPYGCGLLDHVRNRRGLSGQDLNLMCMQIAKMSYLEYDR 840  
QY 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900  
Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPDKWMALESILRRFT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCM 960  
QY 961 IDSECRPRFELVSFSEMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020  
Db 961 IDSECRPRFELVSFSEMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020  
QY 1021 EYILVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEEAPSPAPSEG 1080  
Db 1021 EYILVPOQGFCDPAPGAGGMVHRHRSSTRSGGDLTLGLPSEEEAPSPAPSEG 1080  
QY 1081 AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140  
QY 1141 NQDVRPQPPSPREGPLPAAPAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200  
Db 1141 NQDVRPQPPSPREGPLPAAPAGATLERAKTLPSPKNGVVKDVPFAGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPAENPEYLGLDVPV 1255

RESULT 2  
AAE12130  
ID AAE12130 standard; Protein; 1255 AA.  
XX AC AAE12130;  
XX DT 18-DEC-2001 (first entry)  
XX DE Human tyrosine kinase-type receptor, HER-2.  
XX KW Therapeutic compound; major histocompatibility complex; vaccine;  
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;  
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
KW antigen presenting cell; human; tyrosine kinase-type receptor.  
XX OS Homo sapiens.  
XX FH Key  
XX FT Region  
XX FT 774..782  
XX FT /note= "Antigenic epitope"  
XX PN WO200168677-A2.  
XX PD 20-SEP-2001.  
XX PF 16-MAR-2001; 2001WO-US40328.  
XX PF 16-MAR-2000; 2000US-0527487.  
XX PR (GENZ ) GENZYME CORP.  
XX PA Nicolette CA;  
XX PI  
XX DR WPI; 2001-616284/71.  
XX DR N-PSDB; AAD19731.  
XX PT Novel synthetic therapeutic compound for inducing immune response and  
PT for use in adoptive immunotherapy, has enhanced binding to major  
PT histocompatibility molecules and enhanced immunoregulatory properties

XX  
PS  
XX  
CC The invention relates to synthetic therapeutic compounds (antigenic  
CC peptides) with enhanced binding to major histocompatibility complex  
CC (MHC) molecules and enhanced immunoregulatory properties relative  
CC to their natural counterparts. Compounds of the invention are useful  
CC for inducing an immune response in a subject and for use in adoptive  
CC immunotherapy. They are useful as components of anti-cancer vaccines  
CC and to expand immune effector cells that are specific for cancers  
CC characterised by expression of the breast cancer antigen, HER-2.  
CC Polynucleotides that encode peptides of the invention are useful as  
CC hybridisation probes and as primers for the detection of genes of gene  
CC transcripts that are expressed in antigen presenting cells (APCs), to  
CC confirm transduction of polynucleotides into host cells. The present  
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
CC of the invention are designed based on the HER-2 antigenic peptide  
CC (774-782).  
XX  
SQ Sequence 1255 AA;  
Query Match 97.7%; Score 6659; DB 22; Length 1255;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 1227; Conservative 7; Mismatches 21; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRHLYQGCVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLIQVLPOLCYODTILKDKIFHKNOLA 180  
Db 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGVLIQVLPOLCYODTILKDKIFHKNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKSGRSGESSEDCQSLTRTVTCAGGCARCKPLPTDCCHEQC 240  
Db 181 LTLIDNRSRACHPCSPMCKSGRSGESSEDCQSLTRTVTCAGGCARCKPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHANOEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIT 360  
Db 301 YNYLSTDVGSCTLVCPHANOEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIT 360  
QY 361 ELEFAGCKKI FGSLAFLPESFDGDPASNTAPQVIKANSKFITGLTYLISAWPDSL 420  
Db 361 IOEFAGCKKI FGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEITGLYISAWPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRLRELSGLALIHNTLHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWGLRSRLRELSGLALIHNTLHLCFVHTV 480  
QY 481 PWDQLFRPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540  
Db 481 PWDQLFRPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPGPTQVCNCSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFCVABC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFCVABC 600  
QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCHTSCSYVDLDDKGCPCAPQASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCHTSCSYVDLDDKGCPCAPQASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVWFGLIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNOAQNRILKETEL 720  
Db 661 ILLVVVLGVWFGLIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNOAQNRILKETEL 720



QY 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780  
DB 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780  
QY 781 YVSRLLGICLTSTVQLVTQMPYGCILLDVRNRRGLSGDILLNWCQIAKGSYLEDVR 840  
DB 781 YVSRLLGICLTSTVQLVTQMPYGCILLDVRNRRGLSGDILLNWCQIAKGSYLEDVR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900  
QY 901 HQSDVMSGYVTWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKCM 960  
DB 901 HQSDVMSGYVTWELMTFGAKPYDGPAREIPDLLEKGERLPPOPICTIDVYMWKCM 960  
QY 961 IDSECRPRELVSEFSRARDPQREVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVA 1020  
DB 961 IDSECRPRELVSEFSRARDPQREVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVA 1020  
QY 1021 EYLVLPQOGFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080  
DB 1021 EYLVLPQOGFFCPDPAPGAGMVHRRHSSTRSGGDLTLGLEPSEERAPRSLAPSEG 1080  
QY 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTDGYPVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSTDGYPVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVAFGGAVENTPEYLTQ 1200  
DB 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVAFGGAVENTPEYLTQ 1200  
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255  
DB 1201 GGAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 3  
ID AAB60167  
XX AAB60167 standard; Protein; 1255 AA.  
AC AAB60167;  
XX  
DT 03-APR-2001 (first entry)  
XX  
DE HER2 transgene plasmid construct encoded protein.  
XX  
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
KW antibody.  
XX  
OS Homo sapiens.  
OS Synthetic.  
PN  
PN WO200100244-A2.  
XX  
PD 04-JAN-2001.  
XX  
PF 23-JUN-2000; 2000WO-US17229.  
XX  
PR 25-JUN-1999; 99US-0141316.  
PR 16-MAR-2000; 2000US-0189844.  
XX  
PA (GETH ) GENENTECH INC.  
XX  
PI Erickson S, Schwall R;  
XX  
DR WPI; 2001-061962/07.  
DR N-PSDB; AAF24297.  
XX  
PT Treating tumors, particularly breast cancers, which overexpress an ErbB  
PT receptor and does not respond to an anti-ErbB antibody, comprises  
PT conjugating the antibody to a maytansinoid -  
XX

PS Example 3; Fig 4; 92pp; English.  
XX  
CC The present invention provides a method of treating cancer by  
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In  
CC particular, the antibody is directed against ErbB2 (also known as HER2  
CC and p185neu). The method is particularly useful in the treatment of  
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,  
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.  
XX  
SQ Sequence 1255 AA;  
  
Query Match 97.7%; Score 6659; DB 22; Length 1255;  
Best Local Similarity 97.8%; Pred. No. 0;  
Matches 1227; Conservative 7; Mismatches 21; Indels 0; Gaps 0;  
  
QY 1 MELAALCRGHILLALLPPGAASQVCTGDMKRLRPASPTHLDMLRHLYQGCQVVOGNL 60  
DB 1 MELAALCRGHILLALLPPGAASQVCTGDMKRLRPASPTHLDMLRHLYQGCQVVOGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVRQVPLQRLRIRVGTQLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVRQVPLQRLRIRVGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLEILKGGVLIQIRNPOLCYQDTILMKDIFHKNQLA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLEILKGGVLIQIRNPOLCYQDTILMKDIFHKNQLA 180  
QY 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTYCAGGCARCKGPLPTDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPCKGSRGWESSEDCQSLTRTYCAGGCARCKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHNOEVTABDGTQRCCKSKPCARVCYGLGMQVIKANSKFIGIT 360  
DB 301 YNYLSTDVGSCTLVCPHNOEVTABDGTQRCCKSKPCARVCYGLGMQVIKANSKFIGIT 360  
QY 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPQVIKANSKFIGITELTYLYISAWDPSLP 420  
DB 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPQVIKANSKFIGITELTYLYISAWDPSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRSLRSLGSLALIHNTILCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRSLRSLGSLALIHNTILCFVHTV 480  
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRQEC 540  
DB 481 PWDQLFRNPQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGPTQCVCNCSQFLRQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600  
DB 541 VEECRVLQGLPREYVNAHCLCHPECPONGSVTCFGEADOCVACAHYKDPFPCVARC 600  
QY 601 PSGVKPDLSPYMPKPPDEBEGACQPCPINCTHSCVDLDDKGCAPABORASPLTISVSAVVG 660  
DB 601 PSGVKPDLSPYMPKPPDEBEGACQPCPINCTHSCVDLDDKGCAPABORASPLTISVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAQMRILKETEL 720  
DB 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOTELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780  
DB 721 RKVKVLGSGAGFTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKIILDEAYVMAGVGP 780  
QY 781 YVSRLLGICLTSTVQLVTQMPYGCILLDVRNRRGLSGDILLNWCQIAKGSYLEDVR 840  
DB 781 YVSRLLGICLTSTVQLVTQMPYGCILLDVRNRRGLSGDILLNWCQIAKGSYLEDVR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMALESILRRFT 900

Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEYHADGGKVPKMMALLESILRRRPT 900  
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960  
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIMVKWM 960  
 QY 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFVRSILLEDGMDLVDA 1020  
 Db 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEIDLGPASPLDSTFVRSILLEDGMDLVDA 1020  
 QY 1021 EYILVPOQGFCCPDAPAGAGMWHHRSSSTRSGGDLTLGLFSEEEAPRSLAPSEG 1080  
 Db 1021 EYILVPOQGFCCPDAPAGAGMWHHRSSSTRSGGDLTLGLFSEEEAPRSLAPSEG 1080  
 QY 1081 AGSDVFDGDLGMAKGLQSLPTHPDSPLOQYSESDPTVPLPSETDGYVAPLTCSPQPYV 1140  
 Db 1081 AGSDVFDGDLGMAKGLQSLPTHPDSPLOQYSESDPTVPLPSETDGYVAPLTCSPQPYV 1140  
 QY 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVFAPFGGAVENPEYLTPO 1200  
 Db 1141 NQPDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVFAPFGGAVENPEYLTPO 1200  
 QY 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255  
 Db 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 4

AAU74545  
 ID AAU74545 standard; Protein; 1255 AA.  
 AC AAU74545;

XX  
 DT 23-APR-2002 (first entry)  
 DE Human HER2 (ErbB2) polypeptide.

XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glial disorder; astrocytoma; glioma; glioblastoma; gliosarcoma;  
 KW glandular disorder; macrophagal disorder; epithelial disorder;  
 KW stromal disorder; blastocoeleic disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.

XX Homo sapiens.

XX US2002001587-A1.

XX 03-JAN-2002.

XX 16-MAR-2001; 2001US-081123.

XX 16-MAR-2000; 2000US-189844P.

XX 05-OCT-2000; 2000US-238327P.

XX (ERIC/) ERICKSON S.

XX (SCHW/) SCHWALL R.

XX (SLIW/) SLIWOWSKI M.

XX Erickson S, Schwall R, Sliwowski M;

XX WPI; 2002-163686/21.

XX N-PSDB; ABK14058.

XX Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal -

XX Example 3; Fig 7; 93pp; English.

XX The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytoma, glioma, glioblastoma, gliosarcoma,  
 CC epithelial, stromal, blastocoeleic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match	Best Local Similarity	Score	DB	Length					
Matches 1227; Conservative	97.7%;	6659;	23;	1255;					
	97.8%;	Pred: No. 0;							
	Matches 1227; Conservative	7;	Mismatches	21; Indels 0; Gaps 0;					
QY 1	MELAALCRWGLLLALLPFGAAS	TQVCTG	TMKLR	LPAS	PETHL	DMRL	HL	YQCVVQGNL	60
Db 1	MELAALCRWGLLLALLPFGAAS	TQVCTG	TMKLR	LPAS	PETHL	DMRL	HL	YQCVVQGNL	60
QY 61	ELTYLPTNASLFLQDIQEVQY	VLI	IAHN	QV	QV	QV	QV	QV	120
Db 61	ELTYLPTNASLFLQDIQEVQY	VLI	IAHN	QV	QV	QV	QV	QV	120
QY 121	DPLNNTTPVTGASPGGLREL	QLS	TEIL	KGV	LI	QI	OR	NPOL	180
Db 121	DPLNNTTPVTGASPGGLREL	QLS	TEIL	KGV	LI	QI <td>OR</td> <td>NPOL</td> <td>180</td>	OR	NPOL	180
QY 181	LTLDITNRSRACHPCSPMK	SGRC	WGES	SEDC	QSL	TR	TV	CAGG	240
Db 181	LTLDITNRSRACHPCSPMK	SGRC	WGES	SEDC	QSL	TR	TV	CAGG	240
QY 241	AAGCTGPKHSCLACLP	HNHSGI	CEL	HCP	AL	VT	YNT	DTF	300
Db 241	AAGCTGPKHSCLACLP	HNHSGI	CEL	HCP	AL	VT	YNT	DTF	300
QY 301	YNYLSTDVGSCTLVCP	LHNQ	SV	TA	ED	GT	OR	CE	360
Db 301	YNYLSTDVGSCTLVCP	LHNQ	SV	TA	ED	GT	OR	CE	360
QY 361	ELEPAGCKIFGSLAF	LPES	FD	GP	AS	NT	AP	QY	420
Db 361	ELEPAGCKIFGSLAF	LPES	FD	GP	AS	NT	AP	QY	420
QY 421	DLSVFQNLQVIRGRIL	HN	GA	YS	LT	QGL	IS	WL	480
Db 421	DLSVFQNLQVIRGRIL	HN	GA	YS	LT	QGL	IS	WL	480
QY 481	PWDOLFENPHOALL	TANR	PE	DE	CV	GE	GL	ACH	540
Db 481	PWDOLFENPHOALL	TANR	PE	DE	CV	GE	GL	ACH	540
QY 541	VEECRVLOGLPREY	VN	AR	HCL	PC	HE	Q	P	600
Db 541	VEECRVLOGLPREY	VN	AR	HCL	PC	HE	Q	P	600
QY 601	PSGVKPDLSYMP	PI	WK	FP	DE	E	G	A	660
Db 601	PSGVKPDLSYMP	PI	WK	FP	DE	E	G	A	660
QY 661	ILLVVLGVVFGIL	IK	ER	Q	K	IR	Y	T	720
Db 661	ILLVVLGVVFGIL	IK	ER	Q	K	IR	Y	T	720
QY 721	RKVKVLGSGAF	GV	TY	KG	IN	IP	VA	I	780
Db 721	RKVKVLGSGAF	GV	TY	KG	IN	IP	VA	I	780
QY 781	YVSRLLIGICLT	ST	VQ	LT	Q	L	M	P	840

Db 781 VVRLGLICTSTVQLVTLMPYGCLLDHRNRRGLRGLSQDILLNMQIAKMSYLEVDVR 840  
QY 841 LVHRLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRPT 900  
Db 841 LVHRLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPDKWMALESILRRPT 900  
QY 901 HQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVTMIMVKCWM 960  
Db 901 HQSDVWSGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVTMIMVKCWM 960  
QY 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMDGLDVA 1020  
Db 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMDGLDVA 1020  
QY 1021 EYLVPQOQFFCPDPAGAGGVHHRSSSTRSGGDLTLGLESEEEAPRSLAPSEG 1080  
Db 1021 EYLVPQOQFFCPDPAGAGGVHHRSSSTRSGGDLTLGLESEEEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140  
QY 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVWKDVFAGGAVENPEYLTPO 1200  
Db 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTLPSPKNGVWKDVFAGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFONLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFONLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 5

AAW01111  
ID AAW01111 standard; Protein; 1255 AA.  
AC AAW01111;  
XX  
XX  
XX 01-JAN-1997 (first entry)  
XX  
XX HER-2/neu protein.  
DE  
XX  
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; immunisation; tumour; vaccine; vector.  
XX  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT Domain 676..1255  
FT /label= Intracellular domain  
FT /note= "claimed domain, useful for immunisation"  
XX  
PN W09630514-A1.  
XX  
XX 03-OCT-1996.  
XX  
XX 28-MAR-1996; 96WO-US01689.  
XX  
XX 31-MAR-1995; 95US-0414417.  
XX  
XX (UNIW ) UNIV WASHINGTON.  
PA Cheever MA, Dieis MB;  
XX  
XX WPI: 1996-455361/45.  
DR N-PSDB; AAT40739.  
XX  
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or  
PT treatment of malignancies with which the HER-2/neu oncogene is  
PT associated  
XX  
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.6%; Score 6653; DB 17; Length 1255;  
Best Local Similarity 97.6%; Pred: No. 0;  
Matches 1255; Conservative 8; Mismatches 22; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPFGAASQVCTCTDMKRLPASPETHLDMRLHLYQGCVVQGNL 60  
Db 1 MELAALCRWGLLLALLPFGAASQVCTCTDMKRLPASPETHLDMRLHLYQGCVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIANNQVRQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIANNQVRQVPLQRLRIVRGTLQDPEDNYALAVLDNG 120  
QY 121 DPLNNTTPTVGTASPGGLREQLRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180  
Db 121 DPLNNTTPTVGTASPGGLREQLRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180  
QY 181 LTLIDTNRSRACHPCSPMKSGSRWGSESDQSLTRTVACGACARCKGPLPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKSGSRWGSESDQSLTRTVACGACARCKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTPFESMPNPEGRYTFGASCVTAC 300  
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTYNTDTPFESMPNPEGRYTFGASCVTAC 300  
QY 301 YNLSLTDVSGSTLVCPLHNOEVTAEQTCRCKSPCARVCYGLGMOVIKANSKFIGIT 360  
Db 301 YNLSLTDVSGSTLVCPLHNOEVTAEQTCRCKSPCARVCYGLGMOVIRAVTSAN 360  
QY 361 ELEFAGCKKIFGSLAFPESEFSGDPSASNTAPQYIKANSKFIGITELTYLYISAWPDSL 420  
Db 361 IQEPAGCKKIFGSLAFPESEFSGDPSASNTAPLQPEQLQVFETLEEITGYLYISAWPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLSGSLALIHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGRLSRLSGSLALIHNNHLCFVHTV 480  
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARHCWGPPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHOLCARHCWGPPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVYVNBHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600  
Db 541 VEECRVLQGLPREYVYVNBHCLPCHPECPQNGSVTCFGEADQCVACAHYKPPFCVARC 600  
QY 601 PSQVVPDLSYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPQASPLTSIYSVAVG 660  
Db 601 PSQVVPDLSYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPQASPLTSIYSVAVG 660  
QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRRLQETELVEPLTPSGAMPNQAQRIILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRQOKIRKVTMRRLQETELVEPLTPSGAMPNQAQRIILKETEL 720  
QY 721 RKVKYLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGS 780  
Db 721 RKVKYLGSGAFCTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGS 780  
QY 781 YVSRLLGLICTSTVQLVTLMPYGCLLDHRNRRGLRGLSQDILLNMQIAKMSYLEVDVR 840  
Db 781 YVSRLLGLICTSTVQLVTLMPYGCLLDHRNRRGLRGLSQDILLNMQIAKMSYLEVDVR 840

QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960  
 DB 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960  
 QY 961 IDSECRFRFELVSEFSEMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDGDLVDA 1020  
 DB 961 IDSECRFRFELVSEFSEMRADPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDGDLVDA 1020  
 QY 1021 EYLVPQOGFCPPDPAAGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
 DB 1021 EYLVPQOGFCPPDPAAGAGMVRHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
 QY 1081 AGSDVDFDGLGMAAGLQSLTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
 DB 1081 AGSDVDFDGLGMAAGLQSLTHDPSPQLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
 QY 1141 NOPDVRPQPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVPFAGGAVENPEYLTPO 1200  
 DB 1141 NOPDVRPQPSPREGPLPAARPAATLERAKTILSPGKNGVVKDVPFAGGAVENPEYLTPO 1200  
 QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFGTPTAENPEYLGLDVVP 1255  
 DB 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFGTPTAENPEYLGLDVVP 1255

## RESULT 6

AAW92406  
 ID AAW92406 standard; Protein; 1255 AA.  
 XX AAW92406;  
 AC AAW92406;  
 XX 21-APR-1999 (first entry)  
 DT Human HER-2/neu oncogene protein.  
 XX  
 DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
 XX malignancy; treatment; tumour.  
 KW  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Region 676..1255  
 FT /note= "region which elicits immune response"  
 XX  
 PN US5869445-A.  
 XX  
 XX 09-FEB-1999.  
 XX  
 PF 01-APR-1996; 96US-0625101.  
 XX  
 PR 01-APR-1996; 96US-0625101.  
 PR 17-MAR-1993; 93US-0033644.  
 PR 12-AUG-1993; 93US-0106112.  
 PR 31-MAR-1995; 95US-0414417.  
 XX  
 PA (UNIW ) UNIV WASHINGTON.  
 XX  
 PI Cheever MA, Disis ML;  
 XX  
 DR WPI; 1999-152835/13.  
 DR N-PSDB; AAX01912.  
 XX  
 PT Use of HER-2/neu polypeptides - for eliciting an immune response to  
 PT an HER-2/neu associated malignancy, particularly for treating or  
 PT preventing tumours  
 XX  
 PS Claim 3; Column 31-38; 26pp; English.  
 XX  
 CC This sequence represents the human HER-2/neu oncogene protein. A fragment

of this protein is used in a method for eliciting or enhancing an immune response to HER-2/neu protein. The polypeptide can stimulate T cells and B cells to produce an immune response to the HER-2/neu protein. The method can be used for immunisation against a malignancy in which the HER-2/neu oncogene is associated and in the treatment of an existing tumour, or to prevent tumour occurrence or reoccurrence.

XX Sequence 1255 AA;

Query Match 97.6%; Score 6653; DB 20; Length 1255;  
 Best Local Similarity 97.6%; Pred. No. 0;  
 Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;  
 QY 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHYQSCVQVQGNL 60  
 DB 1 MELAALCRWGLLLALLPPGAASQTCTGDMKRLPASPETHLDMRLHYQSCVQVQGNL 60  
 QY 61 ELTYLPTNASLSFLQDIQEVQYVLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLNG 120  
 DB 61 ELTYLPTNASLSFLQDIQEVQYVLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLNG 120  
 QY 121 DPLNNTTPTVGASPGGLREQLRSLTEILKGGVLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLNG 180  
 DB 121 DPLNNTTPTVGASPGGLREQLRSLTEILKGGVLIQHNQVQVPLQRLRIVRGTLFEDNYALAVLNG 180  
 QY 181 LTLIDTNRSRACHPCSPMKGSRGWCSESSEDQSLTRTVCAAGCARCKGPLPTDCCHQC 240  
 DB 181 LTLIDTNRSRACHPCSPMKGSRGWCSESSEDQSLTRTVCAAGCARCKGPLPTDCCHQC 240  
 QY 241 AGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 DB 241 AGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
 DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
 QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFITGILTYLTVISAWPDSL 420  
 DB 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQEQVFTLEITGILTYLTVISAWPDSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWLGRLSRLGSLGLALHNNTHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLISWLGRLSRLGSLGLALHNNTHLCFVHTV 480  
 QY 481 PWDLFRNPHOALLHTANRPEDECVGGLACHQICARGHCGWGPCTOCVNCQSOFLRGQEC 540  
 DB 481 PWDLFRNPHOALLHTANRPEDECVGGLACHQICARGHCGWGPCTOCVNCQSOFLRGQEC 540  
 QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
 DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
 QY 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCHSCVDLDDKGCQPAEQASPLTISVAVG 660  
 DB 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCHSCVDLDDKGCQPAEQASPLTISVAVG 660  
 QY 661 ILLVVVLGVVFGILIKRRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 DB 661 ILLVVVLGVVFGILIKRRQOKIRKTYMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 QY 721 RYKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780  
 DB 721 RYKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVSP 780  
 QY 781 YVSRLLIGLICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGDOLLNWCQIAKMSYLEDVR 840  
 DB 781 YVSRLLIGLICLTSTVQLVTQMPYGCCLLDHVRNRRGLSGDOLLNWCQIAKMSYLEDVR 840  
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900

QY 901 HQSDVMSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960  
QY 961 IDSECRPRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020  
DB 961 IDSECRPRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020  
QY 1021 EYLYVPOQFFCPDAPGAGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
DB 1021 EYLYVPOQFFCPDAPGAGMVHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
QY 1081 AGSDVFDGLGMAAKGLOSLTPHDPSPLQRYSEDPVPLPSETGYVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGLGMAAKGLOSLTPHDPSPLQRYSEDPVPLPSETGYVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSCKNGVVKDVAFGGAIVENPEYLTQ 1200  
DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLSCKNGVVKDVAFGGAIVENPEYLTQ 1200  
QY 1201 GGAAPQHPHPPAFSFAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255  
DB 1201 GGAAPQHPHPPAFSFAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 7  
ID AAB21198  
XX AAB21198 standard; protein; 1255 AA.  
AC AAB21198;  
DT 12-JAN-2001 (first entry)  
XX Human HER-2/neu protein.  
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX Homo sapiens.  
XX WO200044899-A1.  
XX 03-AUG-2000.  
PF 28-JAN-2000; 2000WO-US02164.  
PR 29-JAN-1999; 99US-0117976.  
PA (CORI-) CORIXA CORP.  
PA (SMIK ) SMITHKLINE BEECHAM.  
PI Cheever MA, Gheysen D;  
DR WPI: 2000-505976/45.  
XX N-PSDB; AAB89736.  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX Claim 52; Fig 7; 128pp; English.  
XX The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.6%; Score 6653; DB 21; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRPASPETHLDMLRHLYQGCVQVGNL 60  
DB 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRPASPETHLDMLRHLYQGCVQVGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPVTGASPGGLRELQRLSRLTEILKGVLIQORNPOLCYQDTILWKDIFHKQNOLA 180  
DB 121 DPLNNTTPVTGASPGGLRELQRLSRLTEILKGVLIQORNPOLCYQDTILWKDIFHKQNOLA 180  
QY 181 LTLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVACGGCARCKGLPTDCCHEOC 240  
DB 181 LTLIDTNRSRACHPCSPMKCKSGRCWGESSEDCQSLTRTVACGGCARCKGLPTDCCHEOC 240  
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTACP 300  
DB 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCYTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHNEVTABDGTORCKSKPCARVCYGLGMQYIKANSKFIGIT 360  
DB 301 YNYLSTDVGSCTLVCPLNHNEVTABDGTORCKSKPCARVCYGLGMHREVRVTSAN 360  
QY 361 ELEFAGCKKIFGSLAPLPESFDGPASNTAPQYIKANSKFIGITELTGVLYSAMDPSLP 420  
DB 361 IOEFAGCKKIFGSLAPLPESFDGPASNTAPQYIKANSKFIGITELTGVLYSAMDPSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNTLHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNTLHLCFVHTV 480  
QY 481 PWDQLFRNPQALLHTANRPEDECYEGGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540  
DB 481 PWDQLFRNPQALLHTANRPEDECYEGGLACHQLCARGHCWPGPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVLOGLPREYVYNARHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPFCVARC 600  
DB 541 VEECRVLOGLPREYVYNARHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPFCVARC 600  
QY 601 PSGVKPDLISYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCPAEORASPLTSIVSAVVG 660  
DB 601 PSGVKPDLISYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCPAEORASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
QY 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780  
DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780  
QY 781 YVSRLLGICLTSTVOLVTQLMPYGLDHRNRRGRGLSGODLLNMCQIAKMSYLEDVR 840  
DB 781 YVSRLLGICLTSTVOLVTQLMPYGLDHRNRRGRGLSGODLLNMCQIAKMSYLEDVR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPPICTIDVYMWKCM 960  
QY 961 IDSECRPRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020

Db 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEDELGPASPLDSTFYRSLLDDDDGLVDA 1020  
Qy 1021 EYLVPOQGFCCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFCCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMGAAGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMGAAGLQSLPTHTDPSPLQRYSDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVPFAFGGAVENPEYLTPO 1200  
Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVPFAFGGAVENPEYLTPO 1200  
Qy 1201 GGAAPOPHPPAFSPAFDNLVYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPOPHPPAFSPAFDNLVYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 8  
AAY84780  
ID AAY84780 standard; Protein; 1255 AA.  
XX  
AC AAY84780;  
XX  
DT 08-AUG-2000 (first entry)  
XX  
DE Amino acid sequence of the SPLICE erbB-2 receptor protein.  
XX  
KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KW bone resorption; inflammatory disease; degenerative disorder;  
KW wound healing.  
XX  
OS Homo sapiens.  
XX  
PN W0200020579-A1.  
XX  
PD 13-APR-2000.  
XX  
PF 01-OCT-1999; 99MO-CA00912.  
XX  
PR 02-OCT-1998; 98US-0165192.  
XX  
PA (UYMC-) UNIV MCMASTER.  
XX  
PI Muller WJ, Siegel PM;  
XX  
DR WPI; 2000-303768/26.  
DR N-PSDB; AAA14812.  
XX  
PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE  
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -  
PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbB-2 receptor protein. The  
XX protein has an in-frame deletion of 16 amino acids, 2 of which are  
XX conserved cysteine residues, compared to the unspliced protein. The  
XX erbB-2 polynucleotide is used to construct probes for detecting  
XX disorders of cell transformation such as cancer. Antibodies to the  
XX protein may be used to detect SPLICE erbB-2 in a sample. Agents  
XX (e.g. antisense oligonucleotides) which inhibit the expression of  
XX SPLICE erbB-2 are useful for reducing tumor cell proliferation and  
XX treating cancer. Substances which stimulate SPLICE erbB-2 are useful  
XX for treating conditions involving damaged cells including conditions  
XX in which degeneration of tissue occurs, such as arthropathy, bone  
XX resorption, inflammatory diseases, degenerative disorders of the  
XX central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 97.6%; Score 6653; DB 21; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGILLALLPPGAASSTOCTCTDMKRLRASPETHLDMRLHYQCCOVQGNL 60  
Db 1 MELAALCRWGILLALLPPGAASSTOCTCTDMKRLRASPETHLDMRLHYQCCOVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLQEDFNEDNALVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHNOVRQVPLQRLRIVRGTLQEDFNEDNALVLDNG 120  
Qy 121 DPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTIILWKDIFHKNNQLA 180  
Db 121 DPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIQRNPQLCYQDTIILWKDIFHKNNQLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSSSEDCQSLSLRTVTCAGGCARCKGPLPTDCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGWSSSEDCQSLSLRTVTCAGGCARCKGPLPTDCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRVTSAN 360  
Qy 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFIGITELTGYLISAWPDSLIP 420  
Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLOPELQVFPETLEEITGYLISAWPDSLIP 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSITLQGLIGISWGLSRLSRELGSLALHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLIGISWGLSRLSRELGSLALHNNHLCFVHTV 480  
Qy 481 PDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCGPGTQCVCNCSQFURGOEC 540  
Db 481 PDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCGPGTQCVCNCSQFURGOEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVARC 600  
Qy 601 PSGVKPDLSTYMPIWKFPDEGACQPCINCTHSCVDLDDKCPAEORASPLTSIISAVVG 660  
Db 601 PSGVKPDLSTYMPIWKFPDEGACQPCINCTHSCVDLDDKCPAEORASPLTSIISAVVG 660  
Qy 661 ILLVVVLGVVFGILLIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGILLIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Qy 781 YVSRLLGICLTSTVOLVTLQMPYGCLLDHVRENRGLSGODLLNKCQIAKMSVLEVR 840  
Db 781 YVSRLLGICLTSTVOLVTLQMPYGCLLDHVRENRGLSGODLLNKCQIAKMSVLEVR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILRRRPT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILRRRPT 900  
Qy 901 HQSDVWSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYIMVVKCM 960  
Db 901 HQSDVWSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDYIMVVKCM 960  
Qy 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEDELGPASPLDSTFYRSLLDDDDGLVDA 1020  
Db 961 IDSECRPRFRELSEFSRMDPQRFVVIQNEDELGPASPLDSTFYRSLLDDDDGLVDA 1020  
Qy 1021 EYLVPOQGFCCPDPAFCAGGMVHRHRSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080

Db	1021	EEYLVVQGGFFCPDPAPGAGMWHRRSSSTRSGGDLTLGLEPSEEARPSPLAPSEG	1080	
Qy	1081	AGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV	1140	
Db	1081	AGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV	1140	
Qy	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVWVDVFAFGGAVENPEYLTQ	1200	
Db	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVWVDVFAFGGAVENPEYLTQ	1200	
Qy	1201	GGAAPQHPHPPAFSFAFNLYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255	
Db	1201	GGAAPQHPHPPAFSFAFNLYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255	
RESULT 9				
AAB85458				
ID	AAB85458 standard; Protein; 1255 AA.			
XX	AAB85458;			
AC				
DT	25-SEP-2001 (first entry)			
XX	Human HER-2/neu protein.			
DE				
XX	Antigen-presenting cell; immunogenic; immune response; HER-2/neu;			
KW	oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.			
XX				
OS	Homo sapiens.			
XX				
PN	WO200153463-A2.			
XX				
PD	26-JUL-2001.			
XX				
PF	19-JAN-2001; 2001WO-US01850.			
XX				
PR	21-JAN-2000; 2000US-0177545.			
XX				
PA	(CORI-) CORIXA CORP.			
XX				
PI	Cheever MA, Hand-Zimmermann S;			
XX				
DR	WPI; 2001-476112/51.			
DR	N-PSDB; AAH23392.			
XX				
PT	New antigen-presenting cells, useful as vaccines for eliciting or			
PT	enhancing an immune response to HER-2/neu protein, particularly useful			
PT	for treating or preventing cancer, e.g. breast cancer -			
XX				
PS	Claim 2; Page 41-46; 49pp; English.			
XX				
CC	The invention provides an isolated antigen-presenting cell, which			
CC	expresses at least an immunogenic portion of a polypeptide that produces			
CC	an immune response to HER-2/neu protein. The antigen-presenting cells are			
CC	useful as vaccines for eliciting or enhancing an immune response to			
CC	HER-2/neu protein, particularly in treating or preventing malignancies in			
CC	which the HER-2/neu oncogene is associated. Specifically, these are			
CC	useful for treating or preventing cancer, e.g. breast cancer, ovarian,			
CC	colon, lung or prostate cancers. The present sequence represents			
CC	the human HER-2/neu protein (also known as p185 or c-erbB2).			
XX				
SQ	Sequence 1255 AA;			
Query Match 97.6%; Score 6653; DB 22; Length 1255;				
Best Local Similarity 97.6%; Pred. No. 0;				
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;				
Qy	1	MELAAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRHLHYQCQVQGNL	60	
Db	1	MELAAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRHLHYQCQVQGNL	60	
Qy	61	ELTYLPTNASLSFLQDIEVQGVLIAHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG	120	

Db	61	ELTYLPTNASLSFLQDIEVQGVLIAHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG	120	
Qy	121	DFLNTTPTVGTASPGRLRLQLRSUTEILKGGVLTQRNPOLCYQDTILWKDIFPHKNOLA	180	
Db	121	DFLNTTPTVGTASPGRLRLQLRSUTEILKGGVLTQRNPOLCYQDTILWKDIFPHKNOLA	180	
Qy	181	LTLIDTNRSRACHPCS PMCKGSRCHGESSEDCQSLTRTVACGACRCKGKPLPTDCCHEOC	240	
Db	181	LTLIDTNRSRACHPCS PMCKGSRCHGESSEDCQSLTRTVACGACRCKGKPLPTDCCHEOC	240	
Qy	241	AAGCTGPKHSDCLACLFHNSGICBLHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300	
Db	241	AAGCTGPKHSDCLACLFHNSGICBLHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP	300	
Qy	301	YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVCYGLGQMWI KANSKFIGIT	360	
Db	301	YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVCYGLGQMEHLREVRVTSAN	360	
Qy	361	ELEFAGCKKI FGS LAF LPES FGD PAS NTAPQYI KANSKFIGITELTGYLTSAMPDLSLP	420	
Db	361	IOEPAGCKKI FGS LAF LPES FGD PAS NTAPLOPEQLQVFEITGYLTSAMPDLSLP	420	
Qy	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRGLALIHNTLHLCFVHTV	480	
Db	421	DLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRGLALIHNTLHLCFVHTV	480	
Qy	481	PHDQLFRPHQALLHTANRPEDECYGBGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC	540	
Db	481	PMDQLFRPHQALLHTANRPEDECYGBGLACHQLCARGHCWGPPTQCVCNCSQFLRGQEC	540	
Qy	541	VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFCGPEADOCVACAHYKOPPCVCARC	600	
Db	541	VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFCGPEADOCVACAHYKOPPCVCARC	600	
Qy	601	PSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG	660	
Db	601	PSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG	660	
Qy	661	ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAOMRILKETEL	720	
Db	661	ILLVVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNQAOMRILKETEL	720	
Qy	721	RKVVLGSGAFGTGVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSP	780	
Db	721	RKVVLGSGAFGTGVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVGSP	780	
Qy	781	YVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRRGLSQDLLNMCQIAKMSYLEVDR	840	
Db	781	YVSRLLGICLTSTVOLVTQLMPYGCLLDHVRENRRGLSQDLLNMCQIAKMSYLEVDR	840	
Qy	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT	900	
Db	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRT	900	
Qy	901	HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPCTIDVTMVMKCMW	960	
Db	901	HQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPCTIDVTMVMKCMW	960	
Qy	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGASPPLDSTFYRSLLEDMDGLVDA	1020	
Db	961	IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGASPPLDSTFYRSLLEDMDGLVDA	1020	
Qy	1021	EYLVPOQGFPCPDPA GAGMVHHRSSSTRSGGDLTLGLEPSEEARPSLAPSEG	1080	
Db	1021	EYLVPOQGFPCPDPA GAGMVHHRSSSTRSGGDLTLGLEPSEEARPSLAPSEG	1080	
Qy	1081	AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV	1140	
Db	1081	AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPTVLPSETDGYVAPLTCSPQPEYV	1140	
Qy	1141	NQPDVVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKNGVWVDVFAFGGAVENPEYLTQ	1200	



Db 1141 NQDVRQPPSPREGPLPAARPAGATLERPKTSLPGKNGVVDVFAFGAVENPEYLTPQ 1200  
AAG88267  
Qy 1201 GGAAPQPHPPAFPAFNDLYWQDPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255  
Db 1201 GGAAPQPHPPAFPAFNDLYWQDPPPERGAPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 10

AAG88267  
ID AAG88267 standard; Protein; 1255 AA.

XX AAG88267;

DT 11-SEP-2001 (first entry)

XX HER2/neu amino acid sequence.

XX Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
XX immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;  
XX tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
XX Homo sapiens.

XX WO200141787-A1.

XX 14-JUN-2001.

XX 11-DEC-2000; 2000WO-US33591.

XX 10-DEC-1999; 99US-0458299.

XX (EPIM-) EPIMUNE INC.

XX Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
XX Keogh E;

XX WPI; 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
XX cellular immune responses for the prevention and treatment of cancer -  
XX Disclosure; Page 15; 199pp; English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).  
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
CC culture in vitro and binds to a complex of an epitope (I), bound to a  
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
CC and a second epitope and the peptide is less than 50 contiguous amino  
CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising  
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic  
CC and immunostimulant activities, and can be used in vaccines. (I), (II)  
CC and (III) are useful for inducing cellular immune responses for the  
CC prevention and treatment of cancer. (I) and (II) are useful for  
CC monitoring or evaluating an immune response to a tumour-associated  
CC antigen when incubated with a T lymphocyte sample from a patient and  
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
CC based vaccines mean that immunosuppressive epitopes that may be present  
CC in whole antigens may be avoided. Selected epitopes may be combined to  
CC enhance immunogenicity. The possible pathological side effects caused by  
CC infectious agents or whole protein antigen is eliminated. The vaccine  
CC provides the ability to direct and focus an immune response to multiple  
CC selected antigens from the same pathogen. Epitope-based anti-tumour  
CC vaccines provides the opportunity to combine epitopes derived from  
CC multiple tumour-associated molecules addressing the problem of tumour-  
CC tumour variability and reducing the likelihood of tumour escape due to  
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
CC the exemplification of the present invention.

XX Sequence 1255 AA;

Query Match 97.6%; Score 6653; DB 22; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLLALLPPGAASSTOCTGTDMLRLPASPEHLDMLRHLQCCQVVOGHL 60  
Db 1 MELAALCRWGLLLALLPPGAASSTOCTGTDMLRLPASPEHLDMLRHLQCCQVVOGHL 60  
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIHQNVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120  
Qy 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGVLIQORNPOLCYQDTILWKDIFHKQNOLA 180  
Db 121 DPLNNTTPTVTCASPGGLRELOLRSLTEILKGVLIQORNPOLCYQDTILWKDIFHKQNOLA 180  
Qy 181 LTLDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVACGACRCKPLPTDCHEOC 240  
Db 181 LTLDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVACGACRCKPLPTDCHEOC 240  
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLYCPLHNOBVTAEQDTCRCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLYCPLHNOBVTAEQDTCRCEKSKPCARVCYGLGMHLEHREVAVTSAN 360  
Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPQYIKANSKFIGITELTGLVLYISAMPDSL 420  
Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEBITGLVLYISAMPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRLSRLGSLALHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRLSRLGSLALHNNHLCFVHTV 480  
Qy 481 PWDQLFRNPHOALLHTANRPEDECEGGLACHOLCARGHGWPGTQCVNCSQFLRGQRC 540  
Db 481 PWDQLFRNPHOALLHTANRPEDECEGGLACHOLCARGHGWPGTQCVNCSQFLRGQRC 540  
Qy 541 VEECRVLQGLPREVYNARHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPFCVARC 600  
Db 541 VEECRVLQGLPREVYNARHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPFCVARC 600  
Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTISVSAVVG 660  
Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTISVSAVVG 660  
Qy 661 ILLVVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGIILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
Qy 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRRGLSQDLNLMQIAKMSYLEVDVR 840  
Db 781 YVSRLLGICLTSTVOLVTLQMPYGLLDHVRNRRGLSQDLNLMQIAKMSYLEVDVR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRRPT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRRPT 900  
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDVYIMVVKCM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPORPCTIDVYIMVVKCM 960  
Qy 961 IDSECRPRFRELNSSEFSRMDARDPQFVVIQNEIDLGPASPLDSTFVRSLLDDDDMDGLVDA 1020  
Db 961 IDSECRPRFRELNSSEFSRMDARDPQFVVIQNEIDLGPASPLDSTFVRSLLDDDDMDGLVDA 1020  
Qy 1021 EYILVPOGPFCDPAPGAGMVHRRSSSTRSGGDLTLGLPEPSEEAAPRSPAPSEG 1080



Db	1021	EEYLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEBAPSLAPSEG	1080
Qy	1081	AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLPGKNGVVKDVFAFGGAVENPEYLTQ	1200
Db	1141	NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLPGKNGVVKDVFAFGGAVENPEYLTQ	1200
Qy	1201	GGAAPQHPHPPAFSFAFNLYWDQDPPPERGAPPSTFKGPTTAENPEYLGLDVVP	1255
Db	1201	GGAAPQHPHPPAFSFAFNLYWDQDPPPERGAPPSTFKGPTTAENPEYLGLDVVP	1255
RESULT 11			
AAE24067			
ID	AAE24067	standard; Protein; 1255 AA.	
XX	XX		
AC	AAE24067;		
XX	XX		
DT	23-SEP-2002	(first entry)	
XX	XX		
DE	Human Her-2 protein.		
XX	XX		
KW	Human; Her-2; epidermal growth factor receptor 2; infection; cancer;		
KW	hyperproliferative disorder; prophylaxis; inflammation; antisease;		
KW	tumour; gene therapy; phosphorothioate backbone.		
XX	XX		
OS	Homo sapiens.		
XX	XX		
PN	WO200222636-A1.		
XX	XX		
PD	21-MAR-2002.		
XX	XX		
PF	12-SEP-2001; 2001WO-US28572.		
XX	XX		
PR	15-SEP-2000; 2000US-0663834.		
XX	XX		
PA	(ISIS-) ISIS PHARM INC.		
XX	XX		
PI	Bennett CF, Cowsert LM;		
XX	XX		
DR	WPI; 2002-471192/50.		
XX	XX		
PT	Novel antisease oligonucleotide which modulates the expression of Human		
PT	Epidermal Growth Factor receptor, Her2, is useful for treating tumors		
PT	Inflammation or to prevent infection in humans -		
XX	XX		
PS	Example 13; Page 95-107; 116pp; English.		
XX	XX		
CC	The invention relates to antisease compounds targetted to a nucleic		
CC	acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)		
CC	that specifically hybridises with and inhibits the expression of Her2.		
CC	Antisease compounds of the invention are used for treating diseases or		
CC	conditions associated with Her2 such as hyperproliferative disorders		
CC	e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,		
CC	neural or cardiac cancer. They are also useful prophylactically e.g.		
CC	to prevent or delay infection, inflammation and tumour formation. The		
CC	invention is also used in gene therapy. The present sequence is human		
CC	Her-2 protein.		
XX	XX		
SQ	Sequence 1255 AA;		
Query Match			
Best Local Similarity 97.6%; Score 6653; DB 23; Length 1255;			
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;			
Qy	1	MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPEHLDMLRHLYQGCQVQGNL	60
Db	1	MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPEHLDMLRHLYQGCQVQGNL	60



Db 1021 EEYLVPOQFFCPDPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLOQRYSEDPVLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSFGKNGVWVDVFAFGGAVENPEYLTPO 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLSFGKNGVWVDVFAFGGAVENPEYLTPO 1200  
Qy 1201 CGAAAPQPPPPAFSDNLYYWDQPPRGGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 CGAAAPQPPPPAFSDNLYYWDQPPRGGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC AAM51143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653

FT Domain /note="extracellular domain"

FT Domain 676..1255

FT Domain /note="intracellular domain"

FT Domain 930..1255

FT Domain /note="phosphorylation domain"

XX WO200212341-A2.

PN 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX or enhancing an immune response to the protein, has Her-2/neu

XX extracellular domain fused to Her-2/neu intracellular or

XX phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX in a variety of cancers, including breast, ovarian, colon, lung and

XX prostate cancer. Her-2/neu is a member of the tyrosine kinase

XX family of receptor-like glycoproteins. It comprises an extracellular

XX domain with homology to the epidermal growth factor receptor

XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

XX intracellular domain that also shows homology to EGFR. Its

XX overexpression correlates with a poor prognosis in breast and

XX ovarian cancers. The invention provides Her-2/neu fusion

XX proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In  
CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
CC protein is fused to a Her-2/neu intracellular domain or  
CC phosphorylation domain (or its DeltapD fragment). An immune  
CC response to Her-2/neu protein is elicited or enhanced by  
CC administering the fusion protein in the form of a vaccine, or by  
CC transfecting cells of an animal ex vivo with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Qy Query Match 97.6%; Score 6653; DB 23; Length 1255;

Best Local Similarity 97.6%; Pred. No. 0;

Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPCAASTQVCTGDMKRLPASPETHLDMRLHYQCCVVGNL 60

Db 1 MELAALCRWGLLLALLPPCAASTQVCTGDMKRLPASPETHLDMRLHYQCCVVGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVQVPLQRLRI VRGTLQFEDNALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVQVPLQRLRI VRGTLQFEDNALAVLDNG 120

Qy 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPQLCYQDTILWKDIFHKKNQLA 180

Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRNPQLCYQDTILWKDIFHKKNQLA 180

Qy 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACRCKGPLPTDCHEQC 240

Db 181 LTLIDTNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGACRCKGPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFPESEFDGDPASNTAPQYIKANSKFIGITELTYGLYISAWPDSL 420

Db 361 IQEFAGCKKIFGSLAFPESEFDGDPASNTAPQYIKANSKFIGITELTYGLYISAWPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGLSLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQFARGQEC 540

Db 481 PWDQLFRPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQFARGQEC 540

Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKPPFPFCVAC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKPPFPFCVAC 600

Qy 601 PSGVKPDLISYMPIWKFPEEGACQPCPINCTHSCVDLDDKGCFAORASPLTSIVSAVVG 660

Db 601 PSGVKPDLISYMPIWKFPEEGACQPCPINCTHSCVDLDDKGCFAORASPLTSIVSAVVG 660

Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLQLQETELVEPLTPSGAMPNQAQMRILKETEL 720

Qy 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGSF 780

Db 721 RKVKVLGSGAGTGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGSF 780

Qy 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGRSGDQLLNWCMQIAKMSYLEVDV 840  
Dd |||||  
Qy 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGRSGDQLLNWCMQIAKMSYLEVDV 840  
Dd |||||  
Qy 841 LVHRDLAARNVLKSPNHVKITDFCLARLLDIDETEHADGGKVPKIMWALESILRRFT 900  
Dd |||||  
Qy 841 LVHRDLAARNVLKSPNHVKITDFCLARLLDIDETEHADGGKVPKIMWALESILRRFT 900  
Dd |||||  
Qy 901 HQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKCM 960  
Dd |||||  
Qy 901 HQSDVWSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYIMVVKCM 960  
Dd |||||  
Qy 961 IDSECRPRFRELVEFSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDDDMDGLVDA 1020  
Dd |||||  
Qy 961 IDSECRPRFRELVEFSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDDDMDGLVDA 1020  
Dd |||||  
Qy 1021 BEYLVPQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Dd |||||  
Qy 1021 BEYLVPQGFPCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Dd |||||  
Qy 1081 AGSDVFDGDLGWAAGKQLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Dd |||||  
Qy 1081 AGSDVFDGDLGWAAGKQLSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Dd |||||  
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVDYPAFGGAVENPEYLTPO 1200  
Dd |||||  
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPFGKNGVVDYPAFGGAVENPEYLTPO 1200  
Dd |||||  
Qy 1201 GGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPU 1255  
Dd |||||  
Qy 1201 GGAAPQPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPU 1255  
Dd |||||

RESULT 14  
AAU77114  
ID AAU77114 standard; Protein; 1255 AA.  
XX  
AC AAU77114;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Human Her-2/neu polypeptide.  
XX  
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;  
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;  
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;  
KW Hodgkin's lymphoma; T cell therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200213847-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 13-AUG-2001; 2001WO-US25408.  
XX  
PR 14-AUG-2000; 2000US-0638280.  
PR 28-SEP-2000; 2000US-0675904.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Gaiger A, Cheever MA, Hand-zimmermann S;  
XX  
DR WPI; 2002-280741/32.  
DR N-PSDB; ABK10730.  
XX  
PT Inhibiting haematological malignancy development by administering  
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide  
PT encoding the polypeptide, or antigen presenting cells expressing the  
PT polypeptide  
XX  
PS Disclosure; Page 71-74; 74pp; English.  
XX

CC The invention relates to a method for inhibiting development of  
CC haematological malignancy in a patient by administering a polypeptide  
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
CC encoding the polypeptide. Antigen presenting cells that express the  
CC protein can also be administered. The sequences are used for inhibiting  
CC development of haematological malignancy such as acute myelogenous  
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
XX  
SQ Sequence 1255 AA;  
  
Query Match 97.6%; Score 6653; DB 23; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;  
  
Qy 1 MELAALCRWGLLLALLPFGAASQVCTGTDMKRLPASPETHLDMLRLHYQCQVQGNL 60  
Dd |||||  
Qy 1 MELAALCRWGLLLALLPFGAASQVCTGTDMKRLPASPETHLDMLRLHYQCQVQGNL 60  
Dd |||||  
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNVQVPLQRLRIVRGTLQFEDNYALAVLNG 120  
Dd |||||  
Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIHNVQVPLQRLRIVRGTLQFEDNYALAVLNG 120  
Dd |||||  
Qy 121 DPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIQPNPQLCYQDTILWKDIFHKQNOLA 180  
Dd |||||  
Qy 121 DPLNNTTPVTGASPGGLREQLRLSLTEILKGGVLIQPNPQLCYQDTILWKDIFHKQNOLA 180  
Dd |||||  
Qy 181 LTLTDTNRSRACHPCSPCKGSCWGESSEDCQSLTRTVACAGGACRCKPLTDCCHEQC 240  
Dd |||||  
Qy 181 LTLTDTNRSRACHPCSPCKGSCWGESSEDCQSLTRTVACAGGACRCKPLTDCCHEQC 240  
Dd |||||  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Dd |||||  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Dd |||||  
Qy 301 YNYLSTDVGSCTLVCPHNVQVTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
Dd |||||  
Qy 301 YNYLSTDVGSCTLVCPHNVQVTAEDGTQRCCKSPCARVCYGLGMEHLREVRAVTSAN 360  
Dd |||||  
Qy 361 ELEFAGCKIFGSLAFIPESFDGDPASNTAPQYIKANSKFITELTGLTYISANPDSL 420  
Dd |||||  
Qy 361 IQEFAGCKIFGSLAFIPESFDGDPASNTAPQYIKANSKFITELTGLTYISANPDSL 420  
Dd |||||  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWGLRSURELSGLALTHNTHLCFVHTV 480  
Dd |||||  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLIGISWGLRSURELSGLALTHNTHLCFVHTV 480  
Dd |||||  
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQCHGCHGPGPTQCVNCSQFLRGQEC 540  
Dd |||||  
Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQCHGCHGPGPTQCVNCSQFLRGQEC 540  
Dd |||||  
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
Dd |||||  
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
Dd |||||  
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAQBARQASPLTSIIVSVVG 660  
Dd |||||  
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAQBARQASPLTSIIVSVVG 660  
Dd |||||  
Qy 661 ILLVVVLGVVFGIILKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Dd |||||  
Qy 661 ILLVVVLGVVFGIILKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
Dd |||||  
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Dd |||||  
Qy 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Dd |||||  
Qy 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGRSGDQLLNWCMQIAKMSYLEVDV 840  
Dd |||||  
Qy 781 YVSRLLGICLTSTVQLTQMPYGCCLLDHVRNRRGRSGDQLLNWCMQIAKMSYLEVDV 840  
Dd |||||  
Qy 841 LVHRDLAARNVLKSPNHVKITDFCLARLLDIDETEHADGGKVPKIMWALESILRRFT 900  
Dd |||||

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900  
QY 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPOLLEKGERLPPOPICTIDVYMIWKCM 960  
Db 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPOLLEKGERLPPOPICTIDVYMIWKCM 960  
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
QY 1021 EYLVPOQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGDLGMAAKGLQSLTPHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEV 1140  
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGAVENPEYLTPO 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSFAFNLYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSFAFNLYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15  
AAR39568  
ID AAR39568 standard; Protein; 1433 AA.  
AC AAR39568;  
XX  
DT 07-FEB-1994 (first entry)  
XX  
DE Sequence of c-erbB-2 tumour antigen.  
XX  
KW Tumour antigen; c-erbB-2; glycoprotein.  
OS Homo sapiens.  
XX  
PN WO9316185-A.  
PD 19-AUG-1993.  
XX  
PF 05-FEB-1993; 93WO-US01055.  
PR 06-FEB-1992; 92US-0831967.  
XX  
PA (CETU ) CETUS ONCOLOGY CORP.  
XX (CREA-) CREATIVE BIOMOLECULES INC.  
PI Houston LL, Huston JS, Oppermann H, Ring DB;  
XX  
DR WPI; 1993-272889/34.  
DR N-PSDB; AAQ46083.  
XX  
PT New single chain Fv polypeptide binding to C-erbB-2 tumour  
XX antigen - for imaging or treating breast or ovarian cancer etc.  
PS Disclosure; pages 48-54; 87pp; English.  
XX  
CC c-erbB-2 refers to a protein antigen expressed on the surface of  
CC tumour cells, such as breast and ovarian tumour cells, which is an  
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents  
CC the location of a stop codon in AAQ46083.  
XX  
SQ Sequence 1433 AA;

Query Match 97.1%; Score 6614; DB 14; Length 1433;  
Best Local Similarity 97.1%; Pred. No. 0;

Matches 1219; Conservative 10; Mismatches 26; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPCCAASQVCTGTDMLKRLUPASPTHLDMLRHLYQCCVQGNL 60  
Db 1 MELAALCRWGLLLALLPCCAASQVCTGTDMLKRLUPASPTHLDMLRHLYQCCVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIEVQGYVLIAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIEVQGYVLIAHNOVRQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPVTGASPGCLRELQRLSLEILKGGVLTORNPOLCYQDTILWKDIFPHKNOLA 180  
Db 121 DPLNNTTPVTGASPGCLRELQRLSLEILKGGVLTORNPOLCYQDTILWKDIFPHKNOLA 180  
QY 181 LTLIDTNRSRACHPCS PMCKGSRGCBESSEDCQSLTRTVACGACRCKGPLPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCS PMCKGSRGCBESSEDCQSLTRTVACGACRCKGPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHNEVTABDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPLNHNEVTABDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360  
QY 361 ELEFAGCKKIFGSLAFLPESFDGPASNTAPOYIKANSKFIGITELTGLYLTISAWPDSL 420  
Db 361 IOEFAGCKKIFGSLAFLPESFDGPASNTAPOYIKANSKFIGITELTGLYLTISAWPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLGSLALIHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLGSLALIHNNTHLSFVHTV 480  
QY 481 PMDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540  
Db 481 PMDQLFRNPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540  
QY 541 VEECEVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600  
Db 541 VEECEVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKDPFCVARC 600  
QY 601 PSGVKPDLISYMPIMKFPDSEEGACQPCINCHTSHCSVDLDDKGPAPORASPLTSIVSAVVG 660  
Db 601 PSGVKPDLISYMPIMKFPDSEEGACQPCINCHTSHCSVDLDDKGPAPORASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVFGILIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYMAGVGGSP 780  
Db 721 RKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKIILDEAYMAGVGGSP 780  
QY 781 YVSRLLGICLTSTVOLVTQLMYPYGLLDHVRNRRGRSGQDLNLMCMQAKGMSYLEDVDR 840  
Db 781 YVSRLLGICLTSTVOLVTQLMYPYGLLDHVRNRRGRSGQDLNLMCMQAKGMSYLEDVDR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRRFT 900  
QY 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPOLLEKGERLPPOPICTIDVYMIWKCM 960  
Db 901 HOSDWSYGVTVWELMTFGAKYDGIIPAREIPOLLEKGERLPPOPICTIDVYMIWKCM 960  
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
QY 1021 EYLVPOQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFPCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGGAAGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRPOPSPREGPLPAARPAGATLERAKTLPFGKNGVVKDVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRPOPSPREGPLPAARPAGATLERPKTLPFGKNGVVKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTTAENPEYLGLDVPV	1255
Db	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 08:41:23  
Job time : 43.9774 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.7573 Seconds

(without alignments)  
4527.811 Million cell updates/sec

Title: SEQ4-632-652-12

Perfect score: 6796

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVFPV 1249

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq\_101002.\*

```

1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6625	97.5	1255	21	Human heregulin 2
2	6625	97.5	1255	22	Human tyrosine kin
3	6625	97.5	1255	22	HER2 transgene pla
4	6625	97.5	1255	23	Human HER2 (ErbB2)
5	6619	97.4	1255	17	HER-2/neu protein.
6	6619	97.4	1255	20	Human HER-2/neu on
7	6619	97.4	1255	21	Amino acid sequenc
8	6619	97.4	1255	21	Human HER-2/neu pr
9	6619	97.4	1255	22	Human HER-2/neu pr
10	6619	97.4	1255	22	HER2/neu amino aci

11	6619	97.4	1255	23	AAE24067	Human Her-2 protei
12	6619	97.4	1255	23	AAE20479	Human Her-2/neu pr
13	6619	97.4	1255	23	AAE11143	Human Her-2/neu po
14	6619	97.4	1255	23	AAU77114	Human Her-2/neu po
15	6576	96.8	1433	14	AAE39568	Sequence of c-erbB
16	6455	95.0	1223	23	AAU98923	Human breast cance
17	6302	92.7	1200	21	AAE21208	Human HER-2/neu pr
18	5861.5	86.2	1256	23	AAE11199	Rat Her-2/neu onco
19	5861.5	86.2	1256	23	AAE11144	Rat Her-2/neu onco
20	5836.5	85.9	1256	21	AAE21206	Mouse Her-2/neu pr
21	5836.5	85.9	1256	21	AAE21206	Mouse Her-2/neu pr
22	5836.5	85.9	1256	22	AAE21206	Amino acid sequenc
23	4817	70.9	919	21	AAE11151	Mouse Her-2/neu on
24	4817	70.9	919	21	AAE11151	Mouse Her-2/neu on
25	4067.5	59.9	920	23	AAE11148	Her-2/neu extracel
26	4067.5	59.9	926	23	AAE11153	Mouse Her-2/neu ex
27	3701	54.5	712	21	AAE21204	Human HER-2/neu fu
28	3701	54.5	712	21	AAE21204	Human HER-2/neu fu
29	3554	52.3	782	18	AAE19764	Her2-GM-CSF immuno
30	3550	52.2	653	21	AAE21200	Extracellular HER-
31	3550	52.2	653	23	AAE11145	Human Her-2/neu on
32	3512	51.7	645	22	AAE60408	Human ErbB2 oncopr
33	3512	51.7	645	22	AAE61593	Human ErbB2 extrac
34	3447	50.7	951	21	AAE44993	DCScFv-erbB2EC fu
35	3344	49.2	624	11	AAE08222	Extracellular port
36	3115	45.8	1210	21	AAE19259	Amino acid sequenc
37	3115	45.8	1210	21	AAE19259	Human EGF receptor
38	3115	45.8	1210	23	AAE23019	Human Her-1 protei
39	3115	45.8	1210	23	AAE23019	Human epidermal gr
40	3113	45.8	1210	22	AAE68420	Amino acid sequenc
41	3085	45.4	583	23	AAE20483	Human protein for
42	3085	45.4	587	23	AAE20481	Human protein for
43	3083	45.4	589	23	AAE20484	Human protein for
44	3083	45.4	600	23	AAE20482	Human protein for
45	3074	45.2	1210	23	ABP51768	Human epidermal gr

#### ALIGNMENTS

```

RESULT 1
AAE92620
ID AAE92620 standard; Protein; 1255 AA.
AC AAE92620;
XX
XX
DT 10-AUG-2000 (first entry)
XX
XX Human heregulin 2 (Her2).
DE
XX
XX Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW self-protein; cancer; breast cancer; prostate cancer;
KW cell-associated peptide antigen; foreign epitope.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 1..173
XX /label= N-terminal
XX /note= "mature polypeptide"
XX Region 5..25
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Region 59..73
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Region 103..117
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Region 149..163
XX /label= insertion region
XX /note= "suitable for foreign epitope insertion"
XX Domain 174..323

```

FT Region /label= Cysteine\_rich\_domain  
210..224  
/label= insertion region  
/note= "suitable for foreign epitope insertion"  
250..264  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
324..483  
FT Domain /label= Ligand\_binding\_domain  
325..339  
/label= insertion region  
/note= "suitable for foreign epitope insertion"  
369..383  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
465..479  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
484..623  
FT Domain /label= Cysteine\_rich\_domain  
579..593  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
624..654  
FT Domain /label= Transmembrane\_domain  
632..652  
/label= insertion region  
/note= "suitable for foreign epitope insertion"  
653..667  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
655..1010  
FT Domain /label= Tyrosine\_kinase\_domain  
661..675  
/label= insertion region  
/note= "suitable for foreign epitope insertion"  
695..709  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
710..730  
FT Region /label= insertion region  
/note= "suitable for foreign epitope insertion"  
1011..1235  
FT Domain /label= C-terminal\_domain  
XX  
W0200020027-A2.  
XX  
13-APR-2000.  
XX  
05-OCT-1999; 99WO-DK00525.  
XX  
05-OCT-1998; 98DK-0001261.  
PR 20-OCT-1998; 98US-0105011.  
XX  
PA (WEBI-) M & E BIOTECH AS.  
XX  
Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
PI Gautam A, Birk P, Karlsson G;  
XX  
WPI; 2000-349917/30.  
DR N-PSDB; AAA09455.  
XX  
Inducing immune responses to weakly immunogenic, tumor associated  
PT peptide antigens for the treatment of breast and prostate cancer  
XX  
PS Claim 62; Page 193-198; 220pp; English.  
XX  
CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
CC response. Subdominant CTL epitopes, antibody binding regions and  
CC cysteine residues involved in disulfide bonds are preserved in the  
CC immunogenised forms. Regions suitable for the insertion of foreign T  
CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
CC cell-associated peptide antigens (PA) such as those associated with  
CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
CC The method comprises effecting simultaneous presentation by antigen  
CC producing cells (APCs) of the animals immune system of: (1) at least 1  
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
CC B-cell group derived from the cell-associated PA; and (2) at least 1  
CC first T helper cell group which is foreign to the animal. Analogues of  
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
CC part of all known and predicted CTL and B-cell epitopes of the respective  
CC PA and including at least one foreign T helper epitope are also claimed.  
CC The method is used to treat prostate, prostate/breast or breast cancer  
CC when the PA is human PSM, FGF8b and Her2, respectively.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6625; DB 21; Length 1255;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;  
QY 1 MELAALCRWGLLLALLPPGAASVCTGTDMLRLPASPETHLMDLRLHLYQGCVVQGNL 60  
DB 1 MELAALCRWGLLLALLPPGAASVCTGTDMLRLPASPETHLMDLRLHLYQGCVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIQEVQGVLLIAHNOVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIQEVQGVLLIAHNOVQVPLQRLRIVRGTOLFEDNYALAVLDNG 120  
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFPHKNOLA 180  
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWKDIFPHKNOLA 180  
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240  
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESNPNGRVTTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESNPNGRVTTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSPCARVCYGLQMVKYKANSKFIGIT 360  
DB 301 YNYLSTDVGSCTLVCPHMQEVTAEQTCRCKSPCARVCYGLQMVKYKANSKFIGIT 360  
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPQLOVFETLEEITGVLYISAWPDSL 420  
DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPQLOVFETLEEITGVLYISAWPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLALIHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRELGLALIHNNTHLCFVHTV 480  
QY 481 PWDQLFRNPQALHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSOFLRGQEC 540  
DB 481 PWDQLFRNPQALHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSOFLRGQEC 540  
QY 541 VEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGPEADQCACAHYKDPFPCVARC 600  
DB 541 VEECRVLOGLPREYVYNAHCLPCHPECPQNGSVTCFGPEADQCACAHYKDPFPCVARC 600  
QY 601 PSGVXPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIQVIKAN 660  
DB 601 PSGVXPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIV----- 655  
QY 661 SKFIGI-----TELKRRQOKIRKYTWRRLLQETELVEPLTPSPGAMPNQAQMRIL 709  
DB 656 SAVVGILLVVLVGVFGIILIKRRQOKIRKYTWRRLLQETELVEPLTPSPGAMPNQAQMRIL 715  
QY 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMA 769  
DB 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMA 775



Qy 770 GVGSPPYVSRLLIGICLTSTVQLVLTQMPYGCCLLDHVRENRLGSGODLLNWCWQIAKGMSY 829  
Db 776 GVGSPPYVSRLLIGICLTSTVQLVLTQMPYGCCLLDHVRENRLGSGODLLNWCWQIAKGMSY 835  
Qy 830 LEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDDIDETEHADGGKVPKMMALESIL 889  
Db 836 LEDVRLVHRDLAARNVLVKSNNHVKITDFGLARLLDDIDETEHADGGKVPKMMALESIL 895  
Qy 890 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLKSGRLPQPPICTIDVYIM 949  
Db 896 RRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLKSGRLPQPPICTIDVYIM 955  
Qy 950 VKCWMIDSECPRELVSEFSRMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDMG 1009  
Db 956 VKCWMIDSECPRELVSEFSRMARDPQRFVWIONEDLGPASPLDSTFYRSLLEDDMG 1015  
Qy 1010 DLVDAEYLVPOQGFPCPDPAAGAGVHHRHRSSTRSGGDLTLGLPSEEEAPRSPL 1069  
Db 1016 DLVDAEYLVPOQGFPCPDPAAGAGVHHRHRSSTRSGGDLTLGLPSEEEAPRSPL 1075  
Qy 1070 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1129  
Db 1076 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135  
Qy 1130 QPEYVNOPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPE 1189  
Db 1136 QPEYVNOPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVFAFGAVENPE 1195  
Qy 1190 YLTPQGGAAQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249  
Db 1196 YLTPQGGAAQPPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 2

AAE12130  
ID AAE12130 standard; Protein; 1255 AA.  
XX  
AC AAE12130;  
XX  
DT 18-DEC-2001 (first entry)  
XX  
DE Human tyrosine kinase-type receptor, HER-2.  
XX  
KW Therapeutic compound; major histocompatibility complex; vaccine;  
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;  
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
KW antigen presenting cell; human; tyrosine kinase-type receptor.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Region 774..782  
FT /note= "Antigenic epitope"  
XX  
XX WO200168677-A2.  
XX  
PD 20-SEP-2001.  
XX  
XX 16-MAR-2001; 2001WO-US40328.  
XX  
XX 16-MAR-2000; 2000US-0527487.  
XX  
XX (GENZ ) GENZYME CORP.  
XX  
XX Nicolette CA;  
XX  
XX WPI; 2001-616284/71.  
DR N-PSDB; AAD19731.  
XX  
XX Novel synthetic therapeutic compound for inducing immune response and  
PT for use in adoptive immunotherapy, has enhanced binding to major  
PT histocompatibility molecules and enhanced immunoregulatory properties

XX  
PS  
XX  
CC The invention relates to synthetic therapeutic compounds (antigenic  
CC peptides) with enhanced binding to major histocompatibility complex  
CC (MHC) molecules and enhanced immunoregulatory properties relative  
CC to their natural counterparts. Compounds of the invention are useful  
CC for inducing an immune response in a subject and for use in adoptive  
CC immunotherapy. They are useful as components of anti-cancer vaccines  
CC and to expand immune effector cells that are specific for cancers  
CC characterised by expression of the breast cancer antigen, HER-2.  
CC Polynucleotides that encode peptides of the invention are useful as  
CC hybridisation probes and as primers for the detection of genes of gene  
CC transcripts that are expressed in antigen presenting cells (APCs), to  
CC confirm transduction of polynucleotides into host cells. The present  
CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
CC of the invention are designed based on the HER-2 antigenic peptide  
CC (774-782).  
XX  
SQ Sequence 1255 AA;  
Query Match 97.5%; Score 6625; DB 22; Length 1255;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;  
Qy 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPTHLDMLRHLVQCQVVGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGDMKRLPASPTHLDMLRHLVQCQVVGNL 60  
Qy 61 ELTYLPTNASLSFLQDQEVQGYVLIHNVQVPLQRLIRVGTQLPEDNYVALVDNG 120  
Db 61 ELTYLPTNASLSFLQDQEVQGYVLIHNVQVPLQRLIRVGTQLPEDNYVALVDNG 120  
Qy 121 DPLNNTTPTVGASPGGLRELQRLSITELTKGVLIOIRNPOLCYQDITLWKDIFHKNNOLA 180  
Db 121 DPLNNTTPTVGASPGGLRELQRLSITELTKGVLIOIRNPOLCYQDITLWKDIFHKNNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVAGGACARCKGPIPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGWESSEDCQSLTRTVAGGACARCKGPIPTDCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACHLPHNHSIGICELHCPALVTYNTDTFESMPNPEGYRTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACHLPHNHSIGICELHCPALVTYNTDTFESMPNPEGYRTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCCKSKPCARVYGLGMOYIKANSKEFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQTCCKSKPCARVYGLGMOYIKANSKEFIGIT 360  
Qy 361 ELEPAGCKKIFGSLAFLEPESFDGDPASNTAPLOEQVQVFTLEITGLYISAMPDLSLP 420  
Db 361 IQEPAGCKKIFGSLAFLEPESFDGDPASNTAPLOEQVQVFTLEITGLYISAMPDLSLP 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGSLGLIHHNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGRLSRLRELGSLGLIHHNTHLCFVHTV 480  
Qy 481 PWDOLFENPHQALLHTANRPEDECVGEGLAGLACHOLCARGHCWGPGTQCVCNQSLRGQEC 540  
Db 481 PWDOLFENPHQALLHTANRPEDECVGEGLAGLACHOLCARGHCWGPGTQCVCNQSLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCVACAHYKDPPEPCVARC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCVACAHYKDPPEPCVARC 600  
Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDDDDKGCPCAEQASPLTSIQYIKAN 660  
Db 601 PSGVKPDLSTYMPIWKFPDEEGACQPCPINCTHSCVDDDDKGCPCAEQASPLTSIV----- 655  
Qy 661 SKFIGI-----TELKRRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRIL 709  
Db 656 SAVVGILLVVLGVVFGILIKRRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQMRIL 715

Claim 4; Page 63-67; 69pp; English.

```
Oy 710 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILD EAYVMA 769
Db 716 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILD EAYVMA 775
Oy 770 GVGSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSY 835
Oy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESIL 895
Oy 890 RRRFTHQSDVMSYGVTVNMLTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIM 949
Db 896 RRRFTHQSDVMSYGVTVNMLTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIM 955
Oy 950 VKCWMIDSECRPRFRELVSERFMRDQRFVVIQNEIDLGPASPLDSTFYRSLLED DDMG 1009
Db 956 VKCWMIDSECRPRFRELVSERFMRDQRFVVIQNEIDLGPASPLDSTFYRSLLED DDMG 1015
Oy 1010 DLVDAEEYLVPOQGFCDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPR SPL 1069
Db 1016 DLVDAEEYLVPOQGFCDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPR SPL 1075
Oy 1070 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
Oy 1130 QPEYVNOQDVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1189
Db 1136 QPEYVNOQDVRPQPSREGPLPAAPAGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1195
Oy 1190 YLTPOGGAAPQHPHPAPFADNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1249
Db 1196 YLTPOGGAAPQHPHPAPFADNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGIDVPV 1255

RESULT 3
ID AAB60167
XX AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
XX
DT 03-APR-2001 (first entry)
XX
DE HER2 transgene plasmid construct encoded protein.
XX
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX
OS Homo sapiens.
XX
XX Synthetic.
XX
XX WO200100244-A2.
XX
XX
XX 04-JAN-2001.
XX
XX 23-JUN-2000; 2000WO-US17229.
XX
XX 25-JUN-1999; 98US-0141316.
XX 16-MAR-2000; 2000US-0189844.
XX
XX (GETH ) GENENTECH INC.
XX
XX Erickson S, Schwall R;
XX
XX WPI: 2001-061962/07.
XX N-PSDB; AAF24297.
XX
XX Treating tumors, particularly breast cancers, which overexpress an ErbB
XX receptor and does not respond to an anti-ErbB antibody, comprises
XX conjugating the antibody to a maytansinoid -
```

PS Example 3; Fig 4; 92pp; English.

CC The present invention provides a method of treating cancer by  
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In  
CC particular, the antibody is directed against ErbB2 (also known as HER2  
CC and p185neu). The method is particularly useful in the treatment of  
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,  
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

SQ Sequence 1255 AA;

Query Match 97.5%; Score 6625; DB 22; Length 1255;  
Best Local Similarity 97.0%; Pred. No. 0;  
Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;

```
Oy 1 MELAALCWGGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLYQCCVQVGNL 60
Db 1 MELAALCWGGLLLALLPPGAASCTGCTDMKRLPASPETHDMLRHLYQCCVQVGNL 60
Oy 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNQVRQVPLQLRIVRGTQLFEDNYALVDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQYVLI AHNQVRQVPLQLRIVRGTQLFEDNYALVDNG 120
Oy 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKQNOLA 180
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGGVLIQRPQLCYQDTILWKDIFHKQNOLA 180
Oy 181 LTLIDTNSRACHPCSPKSGRCWGESSEDCOSLTRTVACGCCARCKGPLTDCCHEOC 240
Db 181 LTLIDTNSRACHPCSPKSGRCWGESSEDCOSLTRTVACGCCARCKGPLTDCCHEOC 240
Oy 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Oy 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCBCKSPCARVCYGLGMQVIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTQRCBCKSPCARVCYGLGMHLEVRVAVTSAN 360
Oy 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLOEQLOVFLEETITGLYIISAWPDSL 420
Db 361 IOEFAGCKKIFGSLAFLPESPDGPASNTAPLOEQLOVFLEETITGLYIISAWPDSL 420
Oy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRELGLALIHNNHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLRELGLALIHNNHLCFVHTV 480
Oy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAHQLCARGHCWGPPTQCVNCSQFLRGQEC 540
Oy 541 VEECRVLOGLPREYVYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKOPPCVARC 600
Db 541 VEECRVLOGLPREYVYVNAHCLPCHPECPONGSVTCFGEADOCVACAHYKOPPCVARC 600
Oy 601 PSGVKPDLSYMPPIWKFPDEBEGACQPCINCHTSVDLDDKGCPEAQRASPLTSIQIKAN 660
Db 601 PSGVKPDLSYMPPIWKFPDEBEGACQPCINCHTSVDLDDKGCPEAQRASPLTSIV---- 655
Oy 661 SKFIGI-----TELKRRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAQMRIL 709
Db 656 SAVVGILLVVLGVVFGILIKRQOKIRKYTNRRLLQETELVEPLTPSGAMPNQAQMRIL 715
Oy 710 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILD EAYVMA 769
Db 716 KETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILD EAYVMA 775
Oy 770 GVGSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSY 829
Db 776 GVGSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNRRGLSGDILLNWCQIAKMSY 835
Oy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESIL 889
```

Db 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHYADGGKVPKIKMALESIL 895  
 Qy 890 RRRFTHQSDVWSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 949  
 Db 896 RRRFTHQSDVWSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955  
 Qy 950 VKCMWIDSECRPRFRELVSERFMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
 Db 956 VKCMWIDSECRPRFRELVSERFMRDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015  
 Qy 1010 DLVDAEYLVPOQGFCCPDPAFGAGVHHRHRSSTRSGGDLTLGLEPSEEAAPRSP 1069  
 Db 1016 DLVDAEYLVPOQGFCCPDPAFGAGVHHRHRSSTRSGGDLTLGLEPSEEAAPRSP 1075  
 Qy 1070 APSEGAGSDVDFDGLGMAAGKQLSLPHTDPSPLQRYSEDPVPLPSETDGVYAPLTCSP 1129  
 Db 1076 APSEGAGSDVDFDGLGMAAGKQLSLPHTDPSPLQRYSEDPVPLPSETDGVYAPLTCSP 1135  
 Qy 1130 OPEYVYNQDVRPQPPSPREGPLPAARPGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1189  
 Db 1136 OPEYVYNQDVRPQPPSPREGPLPAARPGATLERAKTILSPGKNGVVKDVFAGGAVENPE 1195  
 Qy 1190 YLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1249  
 Db 1196 YLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 4

AAU74545  
 ID AAU74545 standard; Protein; 1255 AA.  
 AC AAU74545;

XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human HER2 (ErbB2) polypeptide.  
 XX  
 KW Human; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glial disorder; astrocytoma; glioma; hypothalamic disorder;  
 KW glandular disorder; macropapillary disorder; epithelial disorder;  
 KW stromal disorder; blastocystic disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.

Homo sapiens.

US2002001587-A1.

03-JAN-2002.

16-MAR-2001; 2001US-0811123.

16-MAR-2000; 2000US-189844P.

05-OCT-2000; 2000US-238327P.

(ERIC/) ERICKSON S.

(SCHW/) SCHWALL R.

(SLIW/) SLIWOWSKI M.

Erickson S, Schwall R, Sliwowski M;

WPI; 2002-163686/21.

N-PSDB; ABK14058.

Treating tumour characterised by overexpression of epidermal growth  
 factor receptor, ErbB or cancer in mammal, comprises administering  
 anti-ErbB antibody-maytansinoid conjugate to the mammal

Example 3; Fig 7; 93pp; English.

The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytoma, hypothalamic, glandular, macropapillary,  
 CC epithelial, stromal, blastocystic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.

XX Sequence 1255 AA;

Query Match 97.5%; Score 6625; DB 23; Length 1255;  
 Best Local Similarity 97.0%; Pred. No. 0;

Matches 1222; Conservative 6; Mismatches 16; Indels 16; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASVCTGTDMKRLRLPASPETHLDMRLHLYQCGVQVGNL 60  
 Db 1 MELAALCRWGLLLALLPFGAASVCTGTDMKRLRLPASPETHLDMRLHLYQCGVQVGNL 60  
 Qy 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVQVPLQRLRIVRGTLQDFEDNYALVLDNG 120  
 Db 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHNNQVQVPLQRLRIVRGTLQDFEDNYALVLDNG 120  
 Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGVLTQRPOLCYQDTILWKDILFKNNQLA 180  
 Db 121 DPLNNTTPTVTCASPGGLRELQRLSLEILKGVLTQRPOLCYQDTILWKDILFKNNQLA 180  
 Qy 181 LTLDTNRSRACHPCSPCKSGSRGSESDCQSLTRTVTCAGGACARCKGPLTDCCHEQC 240  
 Db 181 LTLDTNRSRACHPCSPCKSGSRGSESDCQSLTRTVTCAGGACARCKGPLTDCCHEQC 240  
 Qy 241 AAGTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESNPENEGRYTFGASCVTACP 300  
 Db 241 AAGTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESNPENEGRYTFGASCVTACP 300  
 Qy 301 YNYLSTDVGSCTLVCPHNOEVAEDGTQCEKSKPCARVCYGLQVYKANSKFIGIT 360  
 Db 301 YNYLSTDVGSCTLVCPHNOEVAEDGTQCEKSKPCARVCYGLQVYKANSKFIGIT 360  
 Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEITGVLYISAMPDLSL 420  
 Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVFEETLEITGVLYISAMPDLSL 420  
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGI SWLGLRSRLRELGSGLALIHNNTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGI SWLGLRSRLRELGSGLALIHNNTHLCFVHTV 480  
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVNCSQFIRGQBC 540  
 Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCWGPGTQCVNCSQFIRGQBC 540  
 Qy 541 VEECRVLQGLPREYVYNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600  
 Db 541 VEECRVLQGLPREYVYNAHCLPCHPECOPOGNSVTCFGEADQCVACAHYKDPFPCVARC 600  
 Qy 601 PSGVKPDLSPYMPIWKPFDEEGACQPCINCTHSCVDLDDKGCPCAPQASPLTSIYIKAN 660  
 Db 601 PSGVKPDLSPYMPIWKPFDEEGACQPCINCTHSCVDLDDKGCPCAPQASPLTSIYIKAN 660  
 Qy 661 SKFTGI-----TELKROOKIRKYTMRLLOSTELVPELTPSGAMPNQAMRIL 709  
 Db 661 SKFTGI-----TELKROOKIRKYTMRLLOSTELVPELTPSGAMPNQAMRIL 709  
 Qy 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLENTSPKANKILDEAYYMA 769  
 Db 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLENTSPKANKILDEAYYMA 769  
 Qy 716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLENTSPKANKILDEAYYMA 775  
 Db 716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIIPVAIKVLENTSPKANKILDEAYYMA 775  
 Qy 770 GVGSPYVRLILGICLTSTVQLVTQMPYGCILLDVRNRLGSLGSDLLNMCWIAKMSY 829

```
Db 776 GVGSFYVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRRGRIGSQDLLNWCQIAKMSY 835
Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESIL 889
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWMALESIL 895
Qy 890 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIM 949
Db 896 RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTIDVYIM 955
Qy 950 VKCMWIDSECRPRFRELVSERFMRDQRFVVIQNEIDLGPASPLDSTFYRSLLDDDMG 1009
Db 956 VKCMWIDSECRPRFRELVSERFMRDQRFVVIQNEIDLGPASPLDSTFYRSLLDDDMG 1015
Qy 1010 DLVDAEYLVVQOQFFCDDPAPGAGGVHRRSSSTRSGGDLTLGLLEPSEEEAPRSP 1069
Db 1016 DLVDAEYLVVQOQFFCDDPAPGAGGVHRRSSSTRSGGDLTLGLLEPSEEEAPRSP 1075
Qy 1070 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1129
Db 1076 APSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135
Qy 1130 QPEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVXDVFAFGGAVENPE 1189
Db 1136 QPEYVNPDPVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVXDVFAFGGAVENPE 1195
Qy 1190 YLTPOGGAAPQHPHPPAPFAPDNLYYNDODDPERGAPSTFKGTPTAENPEYLGLDVVP 1249
Db 1196 YLTPOGGAAPQHPHPPAPFAPDNLYYNDODDPERGAPSTFKGTPTAENPEYLGLDVVP 1255
```

## RESULT 5

AAW01111  
ID AAW01111 standard; Protein; 1255 AA.

AC AAW01111;

DT 01-JAN-1997 (first entry)

XX HER-2/neu protein.

XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; immunisation; tumour; vaccine; vector.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 676..1255

FT /label= Intracellular domain

FT /note= "Claimed domain, useful for immunisation"

XX WO9630514-A1.

PN 03-OCT-1996.

XX 28-MAR-1996; 96WO-US01689.

XX 31-MAR-1995; 95US-0414417.

XX (UNIW ) UNIV WASHINGTON.

XX Cheever MA, Disie ML;

XX WPI; 1996-455361/45.

XX N-PSDB; AAT40739.

XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or  
PT treatment of malignancies with which the HER-2/neu oncogene is  
PT associated

XX Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transfected host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.

XX Sequence 1255 AA;

Qy Query Match 97.4%; Score 6619; DB 17; Length 1255;

Best Local Similarity 96.9%; Pred No. 0;

Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASVCTCTDMKRLPASPETHLDMLRLHYQSCVVGQNL 60

Db 1 MELAALCRWGLLLALLPPGAASVCTCTDMKRLPASPETHLDMLRLHYQSCVVGQNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLIRVRGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIANQVQVPLQRLIRVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180

Db 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSKGSRGWESSEDCQSLTRTVCAAGCARGKPLPTDCHEQC 240

Db 181 LTLIDTNRSRACHPCSKGSRGWESSEDCQSLTRTVCAAGCARGKPLPTDCHEQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTCASCYTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESNPNEGRTYTCASCYTACP 300

Qy 301 YNYLSTDVGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLCMOVIKANSKFIGIT 360

Db 301 YNYLSTDVGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLCMOHLREVRATVSAN 360

Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFLEBITGYLIYSAMPDLSL 420

Db 361 IQEFAAGCKKIFGSLAFPLPESFDGDPASNTAPLOPEQLQVFLEBITGYLIYSAMPDLSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISHGLRSLRELGSGLALIHNNHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISHGLRSLRELGSGLALIHNNHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPPTQCVNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFFCVARC 600

Db 541 VEECRVLQGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFFCVARC 600

Qy 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPORASPLTSIQIVKAN 660

Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCHTSCVDLDDKGPAPORASPLTSIQIVKAN 660

Qy 661 SKFIGI-----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAOMRIL 709

Db 661 SKFIGI-----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAOMRIL 709

Qy 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMA 769

Db 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDEAYVMA 769

Qy 770 GVGSFYVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRRGRIGSQDLLNWCQIAKMSY 829

Db 770 GVGSFYVSRLLGICLTSTVQLVTLMPYVGLLDHVRNRRGRIGSQDLLNWCQIAKMSY 829

XX Claim 2; Page 56-61; 71pp; English.

QY 830 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889  
 DB 836 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895  
 QY 890 RRRFTHQSDVMSYGVTVNHELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYIM 949  
 DB 896 RRRFTHQSDVMSYGVTVNHELMTFGAKPYDGPAREIPDLLEKGERLPQPPICITIDVYIM 955  
 QY 950 VKCWMIDSECRPRFRELVSERMDARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDMG 1009  
 DB 956 VKCWMIDSECRPRFRELVSERMDARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDMDMG 1015  
 QY 1010 DLVDAEYLVFQOGFCFDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1069  
 DB 1016 DLVDAEYLVFQOGFCFDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSP 1075  
 QY 1070 APSEGAGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1129  
 DB 1076 APSEGAGSDVFDGDLGMAAGLQSLTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP 1135  
 QY 1130 QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPKNGVYKDVFAFGGAVENPE 1189  
 DB 1136 QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTILSPKNGVYKDVFAFGGAVENPE 1195  
 QY 1190 YLTPQ3GAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDVPE 1249  
 DB 1196 YLTPQ3GAAPQHPPPAFSPAFDNLVYWDQPPPERGAPPSTFKGPTAENPEYLGLDVPE 1255  
 RESULT 6  
 AA92406  
 ID AA92406 standard; Protein; 1255 AA.  
 XX  
 AC AA92406;  
 XX  
 DT 21-APR-1999 (first entry)  
 XX  
 DE Human HER-2/neu oncogene protein.  
 XX  
 KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
 KW malignancy; treatment; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 FT Region 676..1255  
 FT /note= "region which elicits immune response"  
 XX  
 XX US5869445-A.  
 XX  
 XX 09-FEB-1999.  
 XX  
 XX 01-APR-1996; 96US-0625101.  
 XX  
 XX 01-APR-1996; 96US-0625101.  
 PR 17-MAR-1993; 93US-0033644.  
 PR 12-AUG-1993; 93US-0106112.  
 PR 31-MAR-1995; 95US-0414417.  
 XX  
 XX (UNIW ) UNIV WASHINGTON.  
 XX  
 XX Cheever MA, Disis ML;  
 XX WPI; 1999-152835/13.  
 DR N-PSDB; AAX01912.  
 XX  
 XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
 PT an HER-2/neu associated malignancy, particularly for treating or  
 PT preventing tumours  
 XX  
 XX Claim 3; Column 31-38; 26pp; English.  
 PS  
 XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune  
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
 CC B cells to produce an immune response to the HER-2/neu protein. The  
 CC method can be used for immunisation against a malignancy in which the  
 CC HER-2/neu oncogene is associated and in the treatment of an existing  
 CC tumour, or to prevent tumour occurrence or reoccurrence.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 97.4%; Score 6619; DB 20; Length 1255;  
 Best Local Similarity 96.9%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;  
 QY 1 MELAALCRWGLLLALLPFGAASCTVCTGDMKRLRLPASPETHDMLRLHYQGCQVVGSL 60  
 DB 1 MELAALCRWGLLLALLPFGAASCTVCTGDMKRLRLPASPETHDMLRLHYQGCQVVGSL 60  
 QY 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
 DB 61 ELTYLPTNASLSFLQDIQEVQYVLIAHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
 QY 121 DPLNNTTPTVTCASPGCLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
 DB 121 DPLNNTTPTVTCASPGCLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKNNQLA 180  
 QY 181 LTLIDTNRSRACHPCSPMKGSRGSESDCQSLTRTVAGGCARCKGPLPTDCHEQC 240  
 DB 181 LTLIDTNRSRACHPCSPMKGSRGSESDCQSLTRTVAGGCARCKGPLPTDCHEQC 240  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPLNHQNVTAEQGTQRCCKSPKPCARVCYGLGMQYIKANSKFEGIT 360  
 DB 301 YNYLSTDVGSCTLVCPLNHQNVTAEQGTQRCCKSPKPCARVCYGLGMHREVRVTSAN 360  
 QY 361 ELEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVFTLEBITGYLYISAMPDSLP 420  
 DB 361 IOEFAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLQVFTLEBITGYLYISAMPDSLP 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLSLQGLGISWLGRLSRLRELGSLALHNNHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLSLQGLGISWLGRLSRLRELGSLALHNNHLCFVHTV 480  
 QY 481 PWDQLFRNPHOALLHTANRDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQBC 540  
 DB 481 PWDQLFRNPHOALLHTANRDECEVGEGLACHQLCARGHCWGPGTQCVCNCSQFLRGQBC 540  
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600  
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPEFCVARC 600  
 QY 601 PSGVKPDLSPYMPIWKFPDEGACPCPINCTHSCVDLDDKGCAPQASPLTSIYIYKAN 660  
 DB 601 PSGVKPDLSPYMPIWKFPDEGACPCPINCTHSCVDLDDKGCAPQASPLTSIYIYKAN 660  
 QY 661 SKFTGI-----TELKROOKIRKYTMRLLOTELVEPLTPSGAMPNQMRIL 709  
 DB 661 SKFTGI-----TELKROOKIRKYTMRLLOTELVEPLTPSGAMPNQMRIL 715  
 QY 710 KETELRKVKVLGSGAFGVYKGIWIPDENVKIPIVAIKVLRNTPSKANKETLDEAYMA 769  
 DB 710 KETELRKVKVLGSGAFGVYKGIWIPDENVKIPIVAIKVLRNTPSKANKETLDEAYMA 769  
 QY 770 GVSGPYVRLIGICLTSTVQLVTQMPYVGLLDHVRNRLGSLQDLNWCWQIAKMSY 829  
 DB 770 GVSGPYVRLIGICLTSTVQLVTQMPYVGLLDHVRNRLGSLQDLNWCWQIAKMSY 835  
 QY 830 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889  
 DB 836 LEDVRLVHRDLAARNVLKSNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895

QY 890 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYIM 949  
| | | | |  
Db 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYIM 955  
| | | | |  
QY 950 VKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
| | | | |  
Db 956 VKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015  
| | | | |  
QY 1010 DLVDAEYLVPOGFFCPDPAAGAGMVHHRSSSTRSGGDLTLGLPSEEEAPRSPL 1069  
| | | | |  
Db 1016 DLVDAEYLVPOGFFCPDPAAGAGMVHHRSSSTRSGGDLTLGLPSEEEAPRSPL 1075  
| | | | |  
QY 1070 APSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPUPSETDGVYAPLTCSP 1129  
| | | | |  
Db 1076 APSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPUPSETDGVYAPLTCSP 1135  
| | | | |  
QY 1130 OBEYNQPDVVRPQPSREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGGAVENPE 1189  
| | | | |  
Db 1136 OBEYNQPDVVRPQPSREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGGAVENPE 1195  
| | | | |  
QY 1190 YLTPQGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPTFKGTPTAENPEYLGLDVVP 1249  
| | | | |  
Db 1196 YLTPQGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPTFKGTPTAENPEYLGLDVVP 1255  
| | | | |

## RESULT 7

AAAB21198  
ID AAB21198 standard; protein; 1255 AA.  
XX AAB21198;  
XX  
DT 12-JAN-2001 (first entry)  
XX Human HER-2/neu protein.  
XX  
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX  
XX Homo sapiens.  
XX WO200044899-A1.  
XX  
XX 03-AUG-2000.  
XX 28-JAN-2000; 2000WO-US021164.  
XX 29-JAN-1999; 99US-0117976.  
XX (CORI-) CORIXA CORP.  
PA (SMIK) SMITHKLINE BEECHAM.  
XX  
PI Cheever MA, Gheysen D;  
XX  
XX WPI: 2000-505976/45.  
DR N-PSDB; AAA89736.  
XX  
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX  
PS Claim 52; Fig 7; 128pp; English.  
XX  
CC The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
XX  
SQ Sequence 1255 AA;  
Query Match 97.4%; Score 6619; DB 21; Length 1255;  
Best Local Similarity 96.9%; Pred. No. 0;  
Matches 1221; Conservative 6; Mismatch 17; Indels 16; Gaps 2;  
QY 1 MELAALCRWGLLLALLPPGAASCTQCTGDMKRLRPASPETHDMLRHLRYQGCVVQGNL 60  
| | | | |  
Db 1 MELAALCRWGLLLALLPPGAASCTQCTGDMKRLRPASPETHDMLRHLRYQGCVVQGNL 60  
| | | | |  
QY 61 ELTYLPTNASSLFLQDIQEVQYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
| | | | |  
Db 61 ELTYLPTNASSLFLQDIQEVQYVLI AHNQVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120  
| | | | |  
QY 121 DLNNTTPTVGASPGGLRELQRLSLEILKGGVLTORNPOLCYODTILWKDIFHKKNOLA 180  
| | | | |  
Db 121 DLNNTTPTVGASPGGLRELQRLSLEILKGGVLTORNPOLCYODTILWKDIFHKKNOLA 180  
| | | | |  
QY 181 LTLIDTNRSRACHPCSPCKGSRGWGSESDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240  
| | | | |  
Db 181 LTLIDTNRSRACHPCSPCKGSRGWGSESDCQSLTRTVCAGGCARCKGPLPTDCCHEQC 240  
| | | | |  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
| | | | |  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
| | | | |  
QY 301 YNYLSTDVGSCTLVCPHNOEVTAE DGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
| | | | |  
Db 301 YNYLSTDVGSCTLVCPHNOEVTAE DGTORCEKSKPCARVCYGLGMOYIRVAVTSAN 360  
| | | | |  
QY 361 ELEFAGCKKIFGSLAFPEFDGDPASNTAPLQPEQLQVFETLEETIGVLYISAWPDSL 420  
| | | | |  
Db 361 IQEAGCKKIFGSLAFPEFDGDPASNTAPLQPEQLQVFETLEETIGVLYISAWPDSL 420  
| | | | |  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNTHLFCFVHT 480  
| | | | |  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNTHLFCFVHT 480  
| | | | |  
QY 481 PWDOLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCNCVSFLRGQEC 540  
| | | | |  
Db 481 PWDOLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQCNCVSFLRGQEC 540  
| | | | |  
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVAHAHKDPFCVVARC 600  
| | | | |  
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVAHAHKDPFCVVARC 600  
| | | | |  
QY 601 PSGVXPDLISYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSIQYIKAN 660  
| | | | |  
Db 601 PSGVXPDLISYMPIWKFPDDEGACQPCPINCTHSCVDLDDKGCPEAQRASPLTSI-----I 655  
| | | | |  
QY 661 SKFIGI-----TELKROOKIRKYTWRRLLQETELVEPLTPSGAMPNQOAMRIL 709  
| | | | |  
Db 656 SAVVGILLVVVLGVVFGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQOAMRIL 715  
| | | | |  
QY 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPPANKEILDEAYVMA 769  
| | | | |  
Db 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPPANKEILDEAYVMA 775  
| | | | |  
QY 770 GVGSPYVSRLLGICLTSTVQLVTQMPYGLLDHVRENRLGSGQDLLNWCQIAKMSY 829  
| | | | |  
Db 776 GVGSPYVSRLLGICLTSTVQLVTQMPYGLLDHVRENRLGSGQDLLNWCQIAKMSY 835  
| | | | |  
QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESIL 889  
| | | | |  
Db 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESIL 895  
| | | | |  
QY 890 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYIM 949  
| | | | |  
Db 896 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPPICTIDVYIM 955  
| | | | |  
QY 950 VKCWMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
| | | | |





Db	1016	DLVDAEYLVPOQFFCPDPAPGCMVHRRHRSSTRSGGDLTLGLEPSEEAAPRSP	1075
Qy	1070	APSEGAGSDVFDGDLGMAAKGLOSLPHDPSPLORYSEDPVPLPSETDGVVAPLTCSP	1129
Db	1076	APSEGAGSDVFDGDLGMAAKGLOSLPHDPSPLORYSEDPVPLPSETDGVVAPLTCSP	1135
Qy	1130	QPEYVYQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKVDFAFGGAVENPE	1189
Db	1136	QPEYVYQPDVVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKVDFAFGGAVENPE	1195
Qy	1190	YLTPOGGAAPQHPHPPAPSPAFDNLYYWDQDPPRGAPSTPKGTPTAENPEYLGDLVPV	1249
Db	1196	YLTPOGGAAPQHPHPPAPSPAFDNLYYWDQDPPRGAPSTPKGTPTAENPEYLGDLVPV	1255

RESULT 9  
ID AAB85458 standard; Protein; 1255 AA.  
XX AAB85458;  
XX 25-SEP-2001 (first entry)  
XX Human HER-2/neu protein.  
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;  
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.  
XX Homo sapiens.  
XX WO200153463-A2.  
XX 26-JUL-2001.  
XX 19-JAN-2001; 2001WO-US01850.  
XX 21-JAN-2000; 2000US-0177545.  
XX (COR1-) CORIXA CORP.  
XX Cheever MA, Hand-Zimmermann S;  
XX WPI; 2001-476112/51.  
XX N-PSDB; AAH23392.  
XX New antigen-presenting cells, useful as vaccines for eliciting or  
PT enhancing an immune response to HER-2/neu protein, particularly useful  
PT for treating or preventing cancer, e.g. breast cancer -  
XX Claim 2; Page 41-46; 49pp; English.  
XX The invention provides an isolated antigen-presenting cell, which  
CC expresses at least an immunogenic portion of a polypeptide that produces  
CC an immune response to HER-2/neu protein. The antigen-presenting cells are  
CC useful as vaccines for eliciting or enhancing an immune response to  
CC HER-2/neu protein, particularly in treating or preventing malignancies in  
CC which the HER-2/neu oncogene is associated. Specifically, these are  
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,  
CC colon, lung or prostate cancers. The present sequence represents  
CC the human HER-2/neu protein (also known as p185 or c-erbB2).  
XX Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 22; Length 1255;  
Best Local Similarity 96.9%; Pred. No. 0;  
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;  
Qy 1 MELAALCRWGLLALLPFGAASTVCTGTDMLKRLPASPETHLMDLRLHYQCQVQGNL 60  
Db 1 MELAALCRWGLLALLPFGAASTVCTGTDMLKRLPASPETHLMDLRLHYQCQVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIEQVGYVLIHNRQVPLQLRLIRVRGTQLPEDNVALAVLNG 120

Db	61	ELTYLPTNASLSFLQDIEQVGYVLIHNRQVPLQLRLIRVRGTQLPEDNVALAVLNG	120
Qy	121	DPLNNTTPTVTCASPGCLRELQRLSLEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA	180
Db	121	DPLNNTTPTVTCASPGCLRELQRLSLEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA	180
Qy	181	LTLIDTNRSRACHPCSPMKSGSRCHGSESDCOSLTRTVVTCAGGCARCKGKPLPTDCHEQC	240
Db	181	LTLIDTNRSRACHPCSPMKSGSRCHGSESDCOSLTRTVVTCAGGCARCKGKPLPTDCHEQC	240
Qy	241	AAGTGPXKHSCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP	300
Db	241	AAGTGPXKHSCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP	300
Qy	301	YNLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMOYIKANSRFIGIT	360
Db	301	YNLSTDVGSCTLVCPHNOEVTAEADGTORCEKSKPCARVCYGLGMEHLREVRVTSAN	360
Qy	361	ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEETGYLYISAWPDSL	420
Db	361	IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFEETLEETGYLYISAWPDSL	420
Qy	421	DLSPFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSRLRELGSLALIHNNTHLCFVHTV	480
Db	421	DLSPFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSRLRELGSLALIHNNTHLCFVHTV	480
Qy	481	PMDQLFRNPQALHTANRPEDECYEGELACHOLCARGHCWGPGTQCNCQFLRGQBC	540
Db	481	PMDQLFRNPQALHTANRPEDECYEGELACHOLCARGHCWGPGTQCNCQFLRGQBC	540
Qy	541	VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVASC	600
Db	541	VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVASC	600
Qy	601	PSGVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDKGPAPQORASPLTSIQYIKAN	660
Db	601	PSGVPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDKGPAPQORASPLTSI-----I	655
Qy	661	SKFIGI-----TELKROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIL	709
Db	656	SAVGILLVWVGVGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIL	715
Qy	710	KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVNA	769
Db	716	KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVNA	775
Qy	770	GVGSPYVSRLLGICLTSTVQLVQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQV	829
Db	776	GVGSPYVSRLLGICLTSTVQLVQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQVLTQV	835
Qy	830	LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPKIMMALESTL	889
Db	836	LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDDIDETEHADGGKVPKIMMALESTL	895
Qy	890	RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIM	949
Db	896	RRRFTHQSDVMSGYVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDVYIM	955
Qy	950	VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG	1009
Db	956	VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG	1015
Qy	1010	DLVDAEYLVPOQFFCPDPAPGAGVHRRHRSSTRSGGDLTLGLEPSEEAAPRSP	1069
Db	1016	DLVDAEYLVPOQFFCPDPAPGAGVHRRHRSSTRSGGDLTLGLEPSEEAAPRSP	1075
Qy	1070	APSEGAGSDVFDGDLGMAAKGLOSLPHDPSPLORYSEDPVPLPSETDGVVAPLTCSP	1129
Db	1076	APSEGAGSDVFDGDLGMAAKGLOSLPHDPSPLORYSEDPVPLPSETDGVVAPLTCSP	1135
Qy	1130	QPEYVYQPDVVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKVDFAFGGAVENPE	1189



Db 1136 QPEYVNPQDVPRPQPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKVFAFGGAVENPE 1195  
QY 1190 YLTPOGGAPOHPHPPAFSPADNLNYYWDQDPPERCAPSTFKGTPTAENPEYGLDVPV 1249  
Db 1196 YLTPOGGAPOHPHPPAFSPADNLNYYWDQDPPERCAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 10  
AAG88267  
ID AAG88267 standard; Protein; 1255 AA.  
XX  
AC AAG88267;  
DT 11-SEP-2001 (first entry)  
XX  
DE HER2/new amino acid sequence.  
XX  
KW Human; HER2/new; epitope; human leukocyte antigen; HLA; T cell;  
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;  
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
XX  
OS Homo sapiens.  
XX  
PN WO200141787-A1.  
PD 14-JUN-2001.  
XX  
PF 11-DEC-2000; 2000WO-US33591.  
XX  
PR 10-DEC-1999; 99US-0458299.  
XX  
PA (EPIM-) EPIMMUNE INC.  
XX  
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Cellis E;  
PI Keogh E;  
XX  
DR WPI; 2001-374995/39.  
XX  
An isolated prepared HER2/new epitope useful in a vaccine for inducing  
cellular immune responses for the prevention and treatment of cancer -  
Disclosure; Page 15; 199pp; English.

The present invention describes isolated prepared HER2/new epitopes (I).  
Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
culture in vitro and binds to a complex of an epitope (I), bound to a  
human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
and a second epitope and the peptide is less than 50 contiguous amino  
acids that have 100% identity with a native peptide sequence of HER2/new;  
(3) a vaccine composition (III) comprising (II) and a pharmaceutical  
excipient; (4) an isolated nucleic acid encoding a peptide comprising  
(I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic  
and immunostimulant activities, and can be used in vaccines. (I), (II)  
and (III) are useful for inducing cellular immune responses for the  
prevention and treatment of cancer. (I) and (II) are useful for  
monitoring or evaluating an immune response to a tumour-associated  
antigen when incubated with a T lymphocyte sample form a patient and  
detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
based vaccines mean that immunosuppressive epitopes that may be present  
in whole antigens may be avoided. Selected epitopes may be combined to  
enhance immunogenicity. The possible pathological side effects caused by  
infectious agents or whole protein antigen is eliminated. The vaccine  
provides the ability to direct and focus an immune response to multiple  
selected antigens from the same pathogen. Epitope-based anti-tumour  
vaccines provides the opportunity to combine epitopes derived from  
multiple tumour-associated molecules addressing the problem of tumour-  
tumour variability and reducing the likelihood of tumour escape due to  
antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
the exemplification of the present invention.

Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 22; Length 1255;

		Best Local Similarity 96.9%; Pred. No. 0;				Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;			
QY	1	MELAAALCRWGLLLALLPPGAAS	TQVCTG	TDMLKRLR	PASPE	THLDMRLH	LYQGQV	VOG	NL 60
Db	1	MELAAALCRWGLLLALLPPGAAS	TQVCTG	TDMLKRLR	PASPE	THLDMRLH	LYQGQV	VOG	NL 60
QY	61	ELTYLPTNASLSFLQDIQEVQ	GVYLI	AHNVQV	PLQRLR	IVRG	TOLF	EDNY	ALAVLDNG 120
Db	61	ELTYLPTNASLSFLQDIQEVQ	GVYLI	AHNVQV	PLQRLR	IVRG	TOLF	EDNY	ALAVLDNG 120
QY	121	DPLNNTTPTVTGASPGGLR	ELQRLS	LEILKGG	VLIQ	RNPOL	CVQD	TLK	DI FHKNNOLA 180
Db	121	DPLNNTTPTVTGASPGGLR	ELQRLS	LEILKGG	VLIQ	RNPOL	CVQD	TLK	DI FHKNNOLA 180
QY	181	LTLIDTNRSRACHPCSP	CMCKSG	RCSG	ESSED	CQS	LT	TR	TV CAGG CARCKG PLPTDC
Db	181	LTLIDTNRSRACHPCSP	CMCKSG	RCSG	ESSED	CQS	LT	TR	TV CAGG CARCKG PLPTDC
QY	241	AAGCTGPKHSDCLAC	LHFNHSGI	CELHCP	ALVTYNT	DTF	FES	MPN	PEGRVTFG
Db	241	AAGCTGPKHSDCLAC	LHFNHSGI	CELHCP	ALVTYNT	DTF	FES	MPN	PEGRVTFG
QY	301	YNYLSTDVGSCTLV	CVPLH	NQEV	TAE	DG	T	OR	CEKSKPCARVCYGL
Db	301	YNYLSTDVGSCTLV	CVPLH	NQEV	TAE	DG	T	OR	CEKSKPCARVCYGL
QY	361	ELEFAGCKKIFGSL	APL	PES	FDG	D	PAS	NTAP	IQPEOLQVF
Db	361	ELEFAGCKKIFGSL	APL	PES	FDG	D	PAS	NTAP	IQPEOLQVF
QY	421	DLVSFQNLQVIR	GRILHNG	AYS	LT	QGL	IS	W	LGLRSLR
Db	421	DLVSFQNLQVIR	GRILHNG	AYS	LT	QGL	IS	W	LGLRSLR
QY	481	PMDQLFRNP	QALHTANR	PE	DE	CV	GE	G	LACHOLCARG
Db	481	PMDQLFRNP	QALHTANR	PE	DE	CV	GE	G	LACHOLCARG
QY	541	VEECRVLOGLPRE	YVYV	NARHCL	PC	HP	EC	Q	PQNGSVTC
Db	541	VEECRVLOGLPRE	YVYV	NARHCL	PC	HP	EC	Q	PQNGSVTC
QY	601	PSGVKPDLSYMP	IWKFP	DE	EG	AC	Q	PC	PIN
Db	601	PSGVKPDLSYMP	IWKFP	DE	EG	AC	Q	PC	PIN
QY	661	SKFIGI	-----	TEL	KR	Q	K	I	R
Db	661	SKFIGI	-----	TEL	KR	Q	K	I	R
QY	710	KETELRKVKVL	SGAF	GVY	K	GI	W	I	P
Db	710	KETELRKVKVL	SGAF	GVY	K	GI	W	I	P
QY	770	GVGSPYVSR	LLGIC	LT	ST	V	Q	L	T
Db	770	GVGSPYVSR	LLGIC	LT	ST	V	Q	L	T
QY	830	LEDVRLVHRD	LAARN	V	L	V	K	S	P
Db	830	LEDVRLVHRD	LAARN	V	L	V	K	S	P
QY	890	RRRTHQSD	V	WSY	G	V	T	W	E
Db	890	RRRTHQSD	V	WSY	G	V	T	W	E
QY	950	VKCMWID	SE	C	R	P	R	E	L
Db	950	VKCMWID	SE	C	R	P	R	E	L
QY	1010	DLVDAEE	V	L	P	O	O	G	F
Db	1010	DLVDAEE	V	L	P	O	O	G	F

Db 1016 DLVDAEYLVPOQGFCDPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPL 1075  
QY 1070 APSEGAGSDVFDGLGMAAGKGLQSLPTHDPSPLQRYSEDPTVLPSETDGVVAPLTCSP 1129  
Db 1076 APSEGAGSDVFDGLGMAAGKGLQSLPTHDPSPLQRYSEDPTVLPSETDGVVAPLTCSP 1135  
QY 1130 QPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVGVKOVFAFGGAVENPE 1189  
Db 1136 QPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVGVKOVFAFGGAVENPE 1195  
QY 1190 YLTPOGGAAPQHPHPPAPSPAFDNLXYNDODPERGAPSTPKGTPTAENPEYLGLDVVP 1249  
Db 1196 YLTPOGGAAPQHPHPPAPSPAFDNLXYNDODPERGAPSTPKGTPTAENPEYLGLDVVP 1255

## RESULT 11

AAE24067  
ID AAE24067 standard; Protein; 1255 AA.

XX AC AAE24067;

XX DT 23-SEP-2002 (first entry)

XX DE Human Her-2 protein.

XX KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;  
KW hyperproliferative disorder; prophylaxis; inflammation; antiseize;  
KW tumour; gene therapy; phosphorothioate backbone.

XX OS Homo sapiens.

XX PN WO200222636-A1.

XX PD 21-MAR-2002.

XX PF 12-SEP-2001; 2001WO-US28572.

XX PR 15-SEP-2000; 2000US-0663834.

XX PA (ISIS-) ISIS PHARM INC.

XX PI Bennett CF, Cowseert LM;

XX DR WPI; 2002-471192/50.

XX DR N-PSDB; AAD38904.

XX PT Novel antiseize oligonucleotide which modulates the expression of Human  
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors  
PT inflammation or to prevent infection in humans -

XX PS Example 13; Page 95-107; 116pp; English.

XX CC The invention relates to antiseize compounds targetted to a nucleic  
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)  
CC that specifically hybridises with and inhibits the expression of Her2.  
CC Antiseize compounds of the invention are used for treating diseases or  
CC conditions associated with Her2 such as hyperproliferative disorders  
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,  
CC neural or cardiac cancer. They are also useful prophylactically e.g.  
CC to prevent or delay infection, inflammation and tumour formation. The  
CC invention is also used in gene therapy. The present sequence is human  
CC Her-2 protein.

XX SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

QY 1 MELALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPEHLDMLRHLVQSCQVVGNNL 60

Db 1 MELALCRWGLLLALLPPGAASCTGCTDMKLRLLPASPEHLDMLRHLVQSCQVVGNNL 60

QY 61 ELTYLPTNASLFLQDIQEVQYVLI AHNVQRPVQLRLRI VRGTQLFEDNTALVALVDNG 120  
Db 61 ELTYLPTNASLFLQDIQEVQYVLI AHNVQRPVQLRLRI VRGTQLFEDNTALVALVDNG 120  
QY 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGGVLIORNPOLCYQDTILWKDIFHKQNOLA 180  
Db 121 DPLNTTPTVGTASPGGLRELQRLSLEILKGGVLIORNPOLCYQDTILWKDIFHKQNOLA 180  
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGACARCKGKPLPTDCCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTRTVACGACARCKGKPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEQGTQRCBCKPCARVCYGLGMEHLREVRVTSAN 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEQGTQRCBCKPCARVCYGLGMEHLREVRVTSAN 360  
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFLEETITGLYISAWPDSLP 420  
Db 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFLEETITGLYISAWPDSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNHLCFVHTV 480  
QY 481 PWDQLFRPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRPHQALLHTANRPEDECYVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFFCVARC 600  
QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSIQIYKAN 660  
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCPAEQRASPLTSI----I 655  
QY 661 SKFIGI-----TELKROQKIRKYTMRRLLOETELVEPLTPSGAMPNQAOMRIL 709  
Db 656 SAVVGILLVVLGVVFGILIKROQKIRKYTMRRLLOETELVEPLTPSGAMPNQAOMRIL 715  
QY 710 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769  
Db 716 KETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 775  
QY 770 GYGSPVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRRGLSGQDLNWCQIAKGMSY 829  
Db 776 GYGSPVSRLLGICLTSTVQLVTLMPYGCCLLDHVRENRRGLSGQDLNWCQIAKGMSY 835  
QY 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889  
Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895  
QY 890 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLUPQPICTIDVYIM 949  
Db 896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLUPQPICTIDVYIM 955  
QY 950 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
Db 956 VKCMWIDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015  
QY 1010 DLVDAEYLVPOQGFCDPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPL 1069  
Db 1016 DLVDAEYLVPOQGFCDPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPL 1075  
QY 1070 APSEGAGSDVFDGLGMAAGKGLQSLPTHDPSPLQRYSEDPTVLPSETDGVVAPLTCSP 1129  
Db 1076 APSEGAGSDVFDGLGMAAGKGLQSLPTHDPSPLQRYSEDPTVLPSETDGVVAPLTCSP 1135  
QY 1130 QPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVGVKOVFAFGGAVENPE 1189  
Db 1136 QPEYVNOQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVGVKOVFAFGGAVENPE 1195  
QY 1190 YLTPOGGAAPQHPHPPAPSPAFDNLXYNDODPERGAPSTPKGTPTAENPEYLGLDVVP 1249  
Db 1196 YLTPOGGAAPQHPHPPAPSPAFDNLXYNDODPERGAPSTPKGTPTAENPEYLGLDVVP 1255

```

Db      :|||||
:136 QPEYVQPDVRPQPSPREGPLAARPAAGATLERPKTUSPGKNGVVKDVFAGGAVENPE 1195
QY      1190 YLTPOGGAAPPHPPAFSPAFDNLNYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249
Db      1196 YLTPOGGAAPPHPPAFSPAFDNLNYYDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 12
ID      AAE20479
AAE20479 standard; Protein; 1255 AA.
AC      XX
AC      AAE20479;
XX      XX
DT      01-JUL-2002 (first entry)
DT      Human Her-2/neu protein.
DE      Human: Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW      human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX      XX
OS      Homo sapiens.
FH      Key
FT      Region
FT      Location/Qualifiers
FT      1021..1030
FT      /note= "Naturally processed HLA-B44-restricted epitope"
XX      WO200214503-A2.
XX      21-FEB-2002.
XX      14-AUG-2001; 2001WO-US41733.
XX      14-AUG-2000; 2000US-225152P.
PR      28-SEP-2000; 2000US-236428P.
PR      21-FEB-2001; 2001US-270520P.
XX      XX
XX      (CORI-) CORIXA CORP.
XX      Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI      McNeill PB, Vedvick TS;
XX      WPI: 2002-280758/32.
XX      N-PSDB: AAD32743.
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT      prevention and diagnosis of cancer, preferably breast cancer
XX      Disclosure; Page 114-117; 129pp; English.
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
CC      effective for eliciting an immune response. The invention is useful for
CC      eliciting an immune response in a patient, where the patient is human
CC      leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
CC      The composition is useful for the therapy and diagnosis of cancer,
CC      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
CC      and other compositions for the diagnosis, prevention and treatment of
CC      human malignancies, for stimulating and/or expanding T cells specific for
CC      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
CC      patient. The invention is useful for stimulating a T cell response in a
CC      human patient, as probe or primer for nucleic acid hybridisation, to
CC      selectively form duplex molecules with complementary stretches of the
CC      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
CC      length gene from a suitable library, and to direct expression of a
CC      polypeptide in appropriate host cells. The composition is useful in
CC      prophylactic or therapeutic applications and for the treatment of cancer,
CC      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
CC      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
XX      Sequence 1255 AA;
XX      Query Match 97.4%; Score 6619; DB 23; Length 1255;

```

```

Best Local Similarity 96.9%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

QY      1 MELAAALCRWGLLLALLPPGAASCTGCTGDMKRLRASPETHLDMRLHLYQCQCVVQGNL 60
Db      1 MELAAALCRWGLLLALLPPGAASCTGCTGDMKRLRASPETHLDMRLHLYQCQCVVQGNL 60

QY      61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASLSFLQDIOEVQGYVLIHAHQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120

QY      121 DPLNNTTPTVTCASPGCLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180
Db      121 DPLNNTTPTVTCASPGCLRELQRLSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180

QY      181 LTLIDTNRSRACHPCSPMKSGSRCHGESSEDCQSLTRTVCAAGCARGKPLPTDCCHEOC 240
Db      181 LTLIDTNRSRACHPCSPMKSGSRCHGESSEDCQSLTRTVCAAGCARGKPLPTDCCHEOC 240

QY      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db      241 AAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300

QY      301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMQVIKANSKFEIGIT 360
Db      301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMQVIKANSKFEIGIT 360

QY      361 ELEFAGCKKIFGSLAFLPESFGDPPASNTAPLOPEQLQVFEETLEBITGYLYISAMPDLSUP 420
Db      361 IQEFAGCKKIFGSLAFLPESFGDPPASNTAPLOPEQLQVFEETLEBITGYLYISAMPDLSUP 420

QY      421 DLSVFQNLQVIRGRILHNGAYSLSLTLQGLIGISWMLGRLSRLRELGSGLALIHNNTHLCFVHTV 480
Db      421 DLSVFQNLQVIRGRILHNGAYSLSLTLQGLIGISWMLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

QY      481 PWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCQFIRGQBC 540
Db      481 PWDQLFRNPHOALLHTANRPEDECVGEGLACHQLCARGHCWGPGTQCVCNCQFIRGQBC 540

QY      541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600
Db      541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPPECVARC 600

QY      601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIQVIKAN 660
Db      601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI-----I 655

QY      661 SKFIGI-----TELKXROOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIL 709
Db      656 SAVVGILLVVVVLGVVVGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQAMRIL 715

QY      710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 769
Db      716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMA 775

QY      770 GVGSPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRLGSDLLNWCNQIAKMSY 829
Db      776 GVGSPYVSRLLGICLTSTVQLVTQMPYGCILLDHVRENRLGSDLLNWCNQIAKMSY 835

QY      830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESTIL 889
Db      836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESTIL 895

QY      890 RRRRTHOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPPOPTCTIDVYIM 949
Db      896 RRRRTHOSDVMSYGVTVWELMTFCAKPYDGIIPAREIPDLLEKGERLPPOPTCTIDVYIM 955

QY      950 VKCWMIDSECRPREFELVSEFSRMARDPQRFVWITQNEDLGPASPLDSTFFYSLLDDDMG 1009
Db      956 VKCWMIDSECRPREFELVSEFSRMARDPQRFVWITQNEDLGPASPLDSTFFYSLLDDDMG 1015

QY      1010 DLVDAEYLVPOQGFCDPDPAGAGVMVHHRSSSTRSGGDLTLTGLEPSEEPASRPL 1069

```

Db 1016 DLVDAEYLVPOQFFCFDPAAGAGMWHHRHRSSTRSGGDLTLGLEPSEEPAPRSPL 1075  
Qy 1070 APSEGAGSDVFDGDLGWAAGLQSLPHDPSPLQRYSEDTPLPSETDGVAPLTCSP 1129  
Db 1076 APSEGAGSDVFDGDLGWAAGLQSLPHDPSPLQRYSEDTPLPSETDGVAPLTCSP 1135  
Qy 1130 QPEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVYVDVFAFGAVENPE 1189  
Db 1136 QPEYVNPQDVVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVYVDVFAFGAVENPE 1195  
Qy 1190 YLTPOGGAOPHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYILGLDVPV 1249  
Db 1196 YLTPOGGAOPHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYILGLDVPV 1255

## RESULT 13

AAM51143

ID AAM51143 standard; Protein; 1255 AA.

XX AC

XX AAM51143;

XX DT 17-JUN-2002 (first entry)

XX DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;  
XX KW tyrosine kinase; receptor; c-erbB2; gene therapy.

XX OS Homo sapiens.

XX FH Key

XX Domain

XX Location/Qualifiers

XX 1..653

XX /note= "extracellular domain"

XX Domain

XX 676..1255

XX /note= "intracellular domain"

XX Domain

XX 990..1255

XX /note= "phosphorylation domain"

XX WO200212341-A2.

XX PN

XX 14-FEB-2002.

XX PD 03-AUG-2001; 2001WO-US24283.

XX PF 03-AUG-2000; 2000US-0632507.

XX PR (CORI-) CORIXA CORP.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX PI Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX DR N-PSDB; ABA92250.

XX XX

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX or enhancing an immune response to the protein, has Her-2/neu

XX extracellular domain fused to Her-2/neu intracellular or

XX phosphorylation domain -

XX Claim 68; Fig 7; 141pp; English.

XX PS

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX in a variety of cancers, including breast, ovarian, colon, lung and

XX prostate cancer. Her-2/neu is a member of the tyrosine kinase

XX family of receptor-like glycoproteins. It comprises an extracellular

XX domain with homology to the epidermal growth factor receptor

XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

XX intracellular domain that also shows homology to EGFR. Its

XX overexpression correlates with a poor prognosis in breast and

XX ovarian cancers. The invention provides Her-2/neu fusion

XX proteins, nucleic acids encoding them, viral vectors, and vaccines

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC comprising the fusion proteins or nucleic acid molecules. In  
CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
CC protein is fused to a Her-2/neu intracellular domain or  
CC phosphorylation domain (or its DeltapD fragment). An immune  
CC response to Her-2/neu protein is elicited or enhanced by  
CC administering the fusion protein in the form of a vaccine, or by  
CC transfecting cells of an animal ex vivo with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;

Best Local Similarity 96.9%; Pred. No. 0;

Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGSCVQVQNL 60

Db 1 MELAALCRWGLLLALLPPGAASCTGCTDMKRLPASPETHLDMRLHYGSCVQVQNL 60

Qy 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRVGTQQLFEDNYALAVLNG 120

Db 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVRQVPLQRLRVGTQQLFEDNYALAVLNG 120

Qy 121 DPLNTPPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180

Db 121 DPLNTPPTVTCASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWKDIFHKQNQLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVCAAGCARGKGLPTDCCHQEC 240

Db 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTVCAAGCARGKGLPTDCCHQEC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVIKANSFPGIT 360

Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVIKANSFPGIT 360

Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPELOVFEETLEETITGLYISAMPDSLP 420

Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLOPELOVFEETLEETITGLYISAMPDSLP 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSRLRELGSGLALIHNNTHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCGWGPQTQVCNCSQFLRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGGLACHQICARGHCGWGPQTQVCNCSQFLRGQEC 540

Qy 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFCPEADQCVACAHYKDPPECVARC 600

Db 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFCPEADQCVACAHYKDPPECVARC 600

Qy 601 PSGVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPAQORASPLTSIQYIKAN 660

Db 601 PSGVKPDLSTYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGCAPAQORASPLTSIQYIKAN 660

Qy 661 SKFTGI-----TELKROOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRIL 709

Db 656 SAVVGILLVWVGVVFGILIKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRIL 715

Qy 710 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYYMA 769

Db 716 KETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYYMA 775

QY 770 GVGSPYVSRLLGICLTSTVLTQMPYGCCLLDHVRENHGRGLSGDQLLNWCQIAKMSY 829  
DB 776 GVGSPYVSRLLGICLTSTVLTQMPYGCCLLDHVRENHGRGLSGDQLLNWCQIAKMSY 835  
QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889  
DB 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 895  
QY 890 RRRFTHQSDVMSYGVTVNLMFTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIM 949  
DB 896 RRRFTHQSDVMSYGVTVNLMFTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIM 955  
QY 950 VKCMIDSECPREFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
DB 956 VKCMIDSECPREFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015  
QY 1010 DLVDAEYLVPOQGFCCPDPAAGAGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1069  
DB 1016 DLVDAEYLVPOQGFCCPDPAAGAGVHHRHRSSTRSGGDLTLGLEPSEEEAPRSP 1075  
QY 1070 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1129  
DB 1076 APSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1135  
QY 1130 QPEYVNOPDVROPPSPREGPLPAARPAGATLERAKTILSPGKNGVYKDVFAFGGAVENTPE 1189  
DB 1136 QPEYVNOPDVROPPSPREGPLPAARPAGATLERAKTILSPGKNGVYKDVFAFGGAVENTPE 1195  
QY 1190 YLTPOGGAAPPHPPAPFSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVPE 1249  
DB 1196 YLTPOGGAAPPHPPAPFSPAFDNLVYWDQPPPERGAPPSTFKGTPPTAENPEYLGLDVPE 1255

## RESULT 14

AAU77114  
ID AAU77114 standard; Protein; 1255 AA.

XX AC AAU77114;

XX DT 05-JUN-2002 (first entry)

XX DE Human Her-2/neu polypeptide.

XX KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;

XX KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;

XX KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;

XX KW Hodgkin's lymphoma; T cell therapy.

XX OS Homo sapiens.

XX PN WO200213847-A2.

XX PD 21-FEB-2002.

XX PF 13-AUG-2001; 2001WO-US25408.

XX PR 14-AUG-2000; 2000US-0638280.

XX PR 28-SEP-2000; 2000US-0675904.

XX PA (CORI-) CORIXA CORP.

XX PI Gaiger A, Cheever MA, Hand-zimmermann S;

XX DR WPI; 2002-280741/32.

XX DR N-PSDB; ABK10730.

XX PT Inhibiting haematological malignancy development by administering

XX PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide

XX PT encoding the polypeptide, or antigen presenting cells expressing the

XX PS polypeptide

XX PS Disclosure; Page 71-74; 74pp; English.

XX PS

XX

CC The invention relates to a method for inhibiting development of  
CC haematological malignancy in a patient by administering a polypeptide  
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
CC encoding the polypeptide. Antigen presenting cells that express the  
CC protein can also be administered. The sequences are used for inhibiting  
CC development of haematological malignancy such as acute myelogenous  
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
XX

SQ Sequence 1255 AA;

Query Match 97.4%; Score 6619; DB 23; Length 1255;  
Best Local Similarity 96.9%; Pred. No. 0;  
Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

QY 1 MELAALCRKWLGLLALLPPGAASVCTCTDKMLRLPASFPETHLDMRLHYQCQVQVGNL 60

DB 1 MELAALCRKWLGLLALLPPGAASVCTCTDKMLRLPASFPETHLDMRLHYQCQVQVGNL 60

QY 61 ELTYLPTNASTLSFLQDIOEVQYVLIHNOVROVPLQRLIRVRGTQLPEDNVALAVLDNG 120

DB 61 ELTYLPTNASTLSFLQDIOEVQYVLIHNOVROVPLQRLIRVRGTQLPEDNVALAVLDNG 120

QY 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

DB 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180

QY 181 LTLIDTNRSRACHPCSPCKGSRGWSSSDQSLTRTVACGGCARGPLPTDCHEQC 240

DB 181 LTLIDTNRSRACHPCSPCKGSRGWSSSDQSLTRTVACGGCARGPLPTDCHEQC 240

QY 241 AAGCTGPKHSDCLACLNHNSGICELHCPALVYNTDTFESMPNPEGYTTCASCVTACP 300

DB 241 AAGCTGPKHSDCLACLNHNSGICELHCPALVYNTDTFESMPNPEGYTTCASCVTACP 300

QY 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMVHREVRVTSAN 360

DB 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMVHREVRVTSAN 360

QY 361 ELEFAGCKKIFGSLAFLPESPDGPASNTAPLQEQLOVFTLEBITGYLIYISAWPDSL 420

DB 361 IQEFAGCKKIFGSLAFLPESPDGPASNTAPLQEQLOVFTLEBITGYLIYISAWPDSL 420

QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLGSRISRELGSGLAIHNNHLCFVHTV 480

DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLISWLGSRISRELGSGLAIHNNHLCFVHTV 480

QY 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHOLCARGHCHGPGTQCVCNCSQFLRGQEC 540

DB 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHOLCARGHCHGPGTQCVCNCSQFLRGQEC 540

QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCACAHYKDPPEVCVARC 600

DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGPEADQCACAHYKDPPEVCVARC 600

QY 601 PSGVKPDLSPYMPYKWFPEDEGACQPCINCHTSCVDLDDKCCPAEORASPLTSIQYIKAN 660

DB 601 PSGVKPDLSPYMPYKWFPEDEGACQPCINCHTSCVDLDDKCCPAEORASPLTSIQYIKAN 660

QY 661 SKFTGI-----TELKRRQOKIRKYTMRLLOBELTELVEPLTPSGAMPNQAOMRIL 709

DB 661 SKFTGI-----TELKRRQOKIRKYTMRLLOBELTELVEPLTPSGAMPNQAOMRIL 709

QY 710 KETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769

DB 710 KETELRKVKVLGSGAFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMA 769

QY 770 GVGSPYVSRLLGICLTSTVLTQMPYGCCLLDHVRENHGRGLSGDQLLNWCQIAKMSY 829

DB 776 GVGSPYVSRLLGICLTSTVLTQMPYGCCLLDHVRENHGRGLSGDQLLNWCQIAKMSY 835

QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIL 889

||||| 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALESIL 895  
890 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIM 949  
896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIM 955  
950 VKCMWIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
956 VKCMWIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015  
1010 DLVDAEYLVPOQGFPCPDPAFCAGMVMHRRSSSTRSGGDLTLGLSPSEEEAPRSP 1069  
1016 DLVDAEYLVPOQGFPCPDPAFCAGMVMHRRSSSTRSGGDLTLGLSPSEEEAPRSP 1075  
1070 APSEGAAGVDFGDLGMAKGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSP 1129  
1076 APSEGAAGVDFGDLGMAKGLQSLPHTDPSLQRYSEDPTVPLPSETDGVVAPLTCSP 1135  
1130 QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVVKOVFAFGAVENPE 1189  
1136 QPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVVKOVFAFGAVENPE 1195  
1190 YLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1249  
1196 YLTPOGGAAPQHPHPPAFSPADNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 15

AAR39568  
ID AAR39568 standard; Protein; 1433 AA.

AC AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

KW Tumour antigen; c-erbB-2; glycoprotein.

OS Homo sapiens.

PN W09316185-A.

PD 19-AUG-1993.

PF 05-FEB-1993; 93WO-US01055.

PR 06-FEB-1992; 92US-0831967.

PA (CETU ) CETUS ONCOLOGY CORP.  
(CREA-) CREATIVE BIOMOLECULES INC.

PI Houston LL, Huston JS, Oppermann H, Ring DB;

XX WPI; 1993-272889/34.

DR N-PSDB; AAQ46083.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour

PT antigen - for imaging or treating breast or ovarian cancer etc.

PS Disclosure; pages 48-54; 87pp; English.

CC c-erbB-2 refers to a protein antigen expressed on the surface of  
CC tumour cells, such as breast and ovarian tumour cells, which is an  
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents  
CC the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 96.8%; Score 6576; DB 14; Length 1433;

Best Local Similarity 96.3%; Pred. No. 0;

Matches 1214; Conservative 8; Mismatches 22; Indels 16; Gaps 2;  
QY 1 MELAALCRWGLLALLPBCAASCTCTDMKRLPASPETHLDMRLHLYQGVVQGNL 60  
DB 1 MELAALCRWGLLALLPBCAASCTCTDMKRLPASPETHLDMRLHLYQGVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVRQVPLQRLRI VRGTQLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNOVRQVPLQRLRI VRGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDITLWKDIPHKNNOLA 180  
DB 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQORNPOLCYQDITLWKDIPHKNNOLA 180  
QY 181 LTLIDTNRSRACHPCSPCKGSRGWSESDCSLTRTVCAGGCARCKGPLPDDCCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPCKGSRGWSESDCSLTRTVCAGGCARCKGPLPDDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSKPCARVCYGLGMOVIKANSKFIGIT 360  
DB 301 YNYLSTDVGSCTLVCPLNHNEVTAEDGTORCEKSKPCARVCYGLGMOVIKANSKFIGIT 360  
QY 361 ELEFAGCKKI FGSLAPLPESPDGPASNTAPLQPELOLVFETLEEITGYLYISAMPDLSL 420  
DB 361 IOEPFAGCKRI FGSLAPLPESPDGPASNTAPLQPELOLVFETLEEITGYLYISAMPDLSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRSLRELGSGLALIHNNTHLCSFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRSLRSLRELGSGLALIHNNTHLCSFVHTV 480  
QY 481 PWDQLFRPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPPTQCVCNSOFLRGQEC 540  
DB 481 PWDQLFRPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPPTQCVCNSOFLRGQEC 540  
QY 541 VEECVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFPCVARC 600  
DB 541 VEECVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHKDPFPCVARC 600  
QY 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSIQIYKAN 660  
DB 601 PSGVKPDLISYMPIWKFPDEEGACQPCINCTHSCVDLDDKGPAPQASPLTSI ----I 665  
QY 661 SKFIGI -----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 709  
DB 661 SKFIGI -----TELKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRIL 715  
QY 710 KETELRKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVMA 769  
DB 716 KETELRKVKVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRNTSPKANKEILDEAYVMA 775  
QY 770 GVGSPVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGRGLSQDLLNCMQIAKMSY 829  
DB 776 GVGSPVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGRGLSQDLLNCMQIAKMSY 835  
QY 830 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALESIL 889  
DB 836 LEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALESIL 895  
QY 890 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIM 949  
DB 896 RRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICITIDVYIM 955  
QY 950 VKCMWIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1009  
DB 956 VKCMWIDSECRPRFRELVSFSEFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMG 1015  
QY 1010 DLVDAEYLVPOQGFPCPDPAFCAGMVMHRRSSSTRSGGDLTLGLSPSEEEAPRSP 1069  
DB 1016 DLVDAEYLVPOQGFPCPDPAFCAGMVMHRRSSSTRSGGDLTLGLSPSEEEAPRSP 1075

Qy	1070	APSEGAGSDVFDGLGMGAAGLSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP	1129
Db	1076	APSEGAGSDVFDGLGMGAAGLSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSP	1135
Qy	1130	OPEYVNPQDVRRQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGGAVENPE	1189
Db	1136	OPEYVNPQDVRRQPPSPREGPLPAARPAAGATLERPKTILSPGKNGVVKDVFAFGGAVENPE	1195
Qy	1190	YLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1249
Db	1196	YLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV	1255

Search completed: July 22, 2003, 09:16:23  
Job time : 42.7573 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.6008 Seconds  
(without alignments)  
5522.503 Million cell updates/sec

Title: SEQ4-632-652-12  
Perfect score: 6796  
Sequence: 1 MELALCRWGLLLALLPPGA.....TFKGTPTAENPEVLGLDPV 1249

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 21:\*
- 1: sp\_archaea:\*
  - 2: sp\_bacteria:\*
  - 3: sp\_fungi:\*
  - 4: sp\_human:\*
  - 5: sp\_invertebrate:\*
  - 6: sp\_mammal:\*
  - 7: sp\_mhc:\*
  - 8: sp\_organelle:\*
  - 9: sp\_phage:\*
  - 10: sp\_plant:\*
  - 11: sp\_rodent:\*
  - 12: sp\_virus:\*
  - 13: sp\_vertebrate:\*
  - 14: sp\_unclassified:\*
  - 15: sp\_rvirus:\*
  - 16: sp\_bacteriap:\*
  - 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6124	90.1	1259	6	018735
2	3109	45.7	1209	11	Q9X70
3	3080	45.3	1210	11	Q9EP98
4	2693	39.6	1165	13	Q9VH40
5	2650.5	39.0	1137	13	Q9W6F6
6	2259	33.2	1328	13	P79754
7	1993.5	29.3	1433	5	Q9BIH9
8	1871	27.5	419	4	Q9UK79
9	1739	25.6	367	11	Q8R2X1
10	1697.5	25.0	412	4	Q8WV0
11	1677	24.7	729	15	Q86712
12	1675	24.6	567	15	Q86714
13	1610.5	23.7	962	15	Q84895
14	1602	23.6	545	15	Q85468
15	1506.5	22.2	655	11	Q9WVF5
16	1490.5	21.9	643	11	Q9ERV6

17	1250	18.4	1193	5	Q9Y1X8	Q9Y1X8 ephydatia f
18	1176.5	17.3	1368	5	Q23821	Q23821 caenorhabdi
19	1148	16.9	1717	5	Q26566	Q26566 echistoeoma
20	1126	16.6	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9SE80	Q9SE80 rattus norv
22	942.5	13.9	599	13	Q9PSH2	Q9PSH2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.1	176	11	Q923V5	Q923V5 rattus norv
25	806.5	11.9	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8SZW1 drosophila
27	754.5	11.1	311	13	Q99162	Q99162 xiphophorus
28	744.5	11.0	1362	13	Q9PVZ4	Q9PVZ4 xenopus lae
29	734	10.8	331	4	Q9BUD7	Q9BUD7 homo sapien
30	723	10.6	149	6	Q9BG66	Q9BG66 oryctolagus
31	713	10.5	1671	5	Q9NJV5	Q9NJV5 biomphalari
32	686	10.1	1418	13	Q93457	Q93457 scophthalmu
33	682.5	10.0	1368	13	Q8UW85	Q8UW85 paralichthy
34	661.5	9.7	1369	13	Q8UW86	Q8UW86 paralichthy
35	656	9.7	1358	13	Q7J798	Q7J798 xenopus lae
36	648.5	9.5	1472	5	Q9USA8	Q9USA8 bombyx mori
37	643	9.5	1412	13	Q8UW84	Q8UW84 paralichthy
38	633.5	9.3	1418	13	Q8UW83	Q8UW83 paralichthy
39	619	9.1	1245	13	Q9YGH8	Q9YGH8 scophthalmu
40	617.5	9.1	2144	5	Q9VD94	Q9VD94 drosophila
41	617	9.1	1371	11	Q9QVW4	Q9QVW4 rattus sp.
42	586.5	8.6	1055	11	O54967	O54967 mus musculu
43	586	8.6	1036	4	Q07912	Q07912 homo sapien
44	574.5	8.5	1091	4	Q9UMQ4	Q9UMQ4 homo sapien
45	565.5	8.3	981	15	O92809	O92809 abelson mur

#### ALIGNMENTS

RESULT 1

018735 ID 018735 PRELIMINARY; PRT; 1259 AA.

AC 018735;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)

DE Erbb-2.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_TaxID=9615;

RN [1] TaxID=9615;

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT "cDNA cloning of erbb-2 from canine mammary gland.";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB008451; BAA23127.1; .

DR HSSP; P11362; 1FGK.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR\_L domain.

DR InterPro; IPR000719; Euk\_pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr\_pkinase.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF01030; Recep\_L\_domain; 2.

DR Pfam; PF02757; YLP; 2.

DR ProDom; PD0000001; Euk\_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TyrKc; 1.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.

KW ATP-binding; Transferrase; Tyrosine-protein kinase.

SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;



Query Match 90.1%; Score 6124; DB 6; Length 1259;  
Best Local Similarity 89.5%; Pred. No. 0;  
Matches 1132; Conservative 42; Mismatches 69; Indels 22; Gaps 4;

Qy 1 MELAAACRWGLLLALLPAGAASTOVCTGTDMKRLPASPELTHDMLRHLVGGQVVOGNL 60  
Db 1 MELAAACRWGLLLALLPAGAASTOVCTGTDMKRLPASPELTHDMLRHLVGGQVVOGNL 60

Qy 61 ELTYLPTNASLFLQDIQEOGVYVLIHNOVROVPIORLRIVRGTQLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLFLQDIQEOGVYVLIHNOVROVPIORLRIVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNTTPTVTCASGGRLQLRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNNOLA 180  
Db 121 DPLNNTTPTVTCASGGRLQLRLSLTEILKGGVLIQNPOLCYQDITLWKDIFHKNNOLA 180

Qy 181 LTLIDTNRSRACHPCSPMKGSRGSCWSSSDCQSLTRTVGAGCARCKGBLPTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGSCWSSSDCQSLTRTVGAGCARCKGBLPTDCCHQC 240

Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMYIKANSKFITG 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQCEKSKPCARVCYGLGMYIKANSKFITG 360

Qy 361 ELFPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLBEITGYLYISAWPDSLP 420  
Db 361 IQEPAGCKKIFGSLAFIPESFDGDPASNTAPLQEQLOVFETLBEITGYLYISAWPDSLP 420

Qy 421 DLSVFONLQVIRGRILHNGAYSFLTGLGISWLGSLRLSGSLALIHNTLHLCFVHTV 480  
Db 421 NLSVFONLQVIRGRILHNGAYSFLTGLGISWLGSLRLSGSLALIHNTLHLCFVHTV 480

Qy 481 PWQLFRNPHOALLHTANRDECEVCGELACHQLCARGHCWGPPTQCVNCSQFLRQEC 540  
Db 481 PWQLFRNPHOALLHTANRDECEVCGELACHQLCARGHCWGPPTQCVNCSQFLRQEC 540

Qy 541 VEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCAVAKYDPFCVARC 600  
Db 540 VEECRVLOGLPREVYNARHCLPCHPECPQNGSVTCFGEADQCAVAKYDPFCVARC 600

Qy 601 PSGVKPDLSPYMPIKWFDEEGACOPCPINCTHSCVDLDKGCPAEQRASPLTSIQYIKAN 660  
Db 600 PSGVKPDLSPYMPIKWFDEEGACOPCPINCTHSCVDLDKGCPAEQRASPLTSIQYIKAN 660

Qy 661 SKFIGI-----TELKRRQOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRIL 709  
Db 655 AAVVGILLVAVVGLVLGILIKRRRQOKIRKYTMRELLQETELVEPLTPSGAMPNQAQMRIL 714

Qy 710 KETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYMA 769  
Db 715 KETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLENTSPKANKEILDEAYMA 774

Qy 770 GVGSPPYSRLIGLICLTSTVOLVTQMPYGLCLLDHVRNRRGLSGQDLLNWCMIKAGMSY 829  
Db 775 GVGSPPYSRLIGLICLTSTVOLVTQMPYGLCLLDHVRNRRGLSGQDLLNWCMIKAGMSY 834

Qy 830 LEDVRLVHRDLAARNVLKSPNHYKIIDFGLARLLDIDETEHADGKVPKIMWALESI 889  
Db 835 LEDVRLVHRDLAARNVLKSPNHYKIIDFGLARLLDIDETEHADGKVPKIMWALESI 894

Qy 890 RRRFTHOSDVSYGVTYVWELMTFCAPYDGPAREIPDLLEKGERLPQPPICITDVYIM 949  
Db 895 RRRFTHOSDVSYGVTYVWELMTFCAPYDGPAREIPDLLEKGERLPQPPICITDVYIM 954

Qy 950 VKCWMIDSECRPREFELVSEFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDMG 1009  
Db 955 VKCWMIDSECRPREFELVSEFSRWARDPQRFVVIQNEEDLGPASPLDSTFYRSLLEDDMG 1014

Qy 1010 DLVDAEYLVPQQGFCCPEPTPCAGGTAAHRRHSSSTRNGGELTLGLEPSEEPKSP 1074

Db 1015 DLVDAEYLVPQQGFCCPEPTPCAGGTAAHRRHSSSTRNGGELTLGLEPSEEPKSP 1074

Qy 1070 APSEGAGSDVFDGDLGNGAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1129

Db 1075 APSEGAGSDVFDGDLGNGAAGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSP 1134

Qy 1130 QPYVNVQPDVVRPQPPSPREGPLPAARPAGATTLER-----AKTILSPGKNGVVKDVFARFGA 1184

Db 1135 QPYVNVQPDVVRPQPPSPREGPLPAARPAGATTLER-----AKTILSPGKNGVVKDVFARFGA 1194

Qy 1185 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSPSTFKGTPTAENPEYLG 1244

Db 1195 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQPPPERGAPSPSTFKGTPTAENPEYLG 1254

Qy 1245 LDVPV 1249

Db 1255 LDVPV 1259

RESULT 2  
Q9QX70 PRELIMINARY; PRT; 1209 AA.

AC Q9QX70;  
DT 01-MAY-2000 (T-EMBLrel. 13, Created)  
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)  
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)  
DE Epidermal growth factor receptor.  
GN EGFR.

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OC NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RX MEDLINE=9025888; PubMed=2342466;  
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
RA Earp H.S.;  
RT "A truncated, secreted form of the epidermal growth factor receptor is  
RT encoded by an alternatively spliced transcript in normal rat tissue."  
RL Mol. Cell. Biol. 10:2973-2982(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RA Petch L.A.;  
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FISHER; TISSUE=LIVER;  
RA Guttridge K., Dawson T.L., Earp H.S.;  
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; M37394; AAF14008.1; -.  
DR HSP; P11362; IFGK.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recept\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TYRK; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7F6CC1B7773 CRC64;

Query Match 45.7%; Score 3109; DB 11; Length 1209;  
Best Local Similarity 50.0%; Pred. No. 8.7e-226;







Best Local Similarity 46.8%; Pred. No. 3.5e-191; Matches 533; Conservative 165; Mismatches 358; Indels 83; Gaps 26;	
QY	161 LCYQDTILWKDIFHKQNKALTLIDNRSRACHPCSPMKSGRCWGESSEDCQSLTRTVIC 220
DB	3 LCFADTIHQDIIVRNPWASNFLLVPTNGSGGCRCHKSCGT-RCWGPTEHNHCQTLTKTVIC 61
QY	221 AGGC-ARCKGPLPTDCHCQCAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTF 279
DB	62 AEQDGRGCVGVSDCHREACGCGSPKDTDFACWNFNDSGACVTCQPTQFVYNTTF 121
QY	280 ESNPNPEGYTTFGASCVTACPVNYLSTVGSCTLVCLPHNQEVTAEDGTORCEKCSKPCA 339
DB	122 OLEHNHNAKTYTGAFCKKCPHNFV-VDSSCVACFPSSKMEV-EENGIMCKPCDTCIP 179
QY	340 RUCYGLGMOYIKANSKFIIGITILE-PAGCKKIIGSLAFIPESDFGDPASNTAPLQEQLO 398
DB	180 KACDGGTGS-LVSAQTVDSSNIDKFINCTKINGNLIPLVTHGDPYHTIAINPEKLN 238
QY	399 VFETLEETGYLISAWPDSLPDLSVFQNLQVIRGILHNGAYSLTLQGLGISWLGRLSL 458
DB	239 IFQTVREITGYLNIQSPENMTDFRVFSLVTIGGRALYSGLSLILKQGGITSLQFQSL 298
QY	459 RELGSLAIHHNTHLCFVHTVPWDQLFRNPHQALLHTANRPDECVGEGLAGHQICARG 518
DB	299 KOISAGNIYITDNSLNCYVHTVNTSLFSTPSQKTVIHRNKAENCTADGMVCNELCSSD 358
QY	519 HCWGPPTQVNCVSOFLRGQECVEECRVLOGLPREYVNAHCLPCHPEQCP-QNGSVTCF 577
DB	359 GCWGPDPQCLSKRIRGRTCTIESCNLYDGEFREFANGSVCMEDCPQCEKMDNMITCY 418
QY	578 GPADOCVACAHYKDPFPFCVARGSPGVKPDLSYMPIWKFDPDEGACQPCPINCHSCVDL 637
DB	419 GPGPDHCTCFHFKDGNVCVKCPDGLQANSF--IFKYADEBRECHPCHPNCTQCRGP 476
QY	638 DDKGC-----PAORASPLTSIQYIKANSKFI-----GIT--ELKRRQOKIRKYT 680
DB	477 ASHDCIYYPWTROSTLPOHAR-TPLIAAGVI--GGLFIIIMGLTFAVYVRRKSIKKRA 533
QY	681 MRELLOETELVEPLTSGAMPNQAOMRILKETELRKVKVLGSGAFGVYKGIWIPGENV 740
DB	534 LRFL-ETELVEPLTSGTAPNQAORILKETELRKVKVLGSGAFGVYKGIWIPGETV 592
QY	741 KIPVAIKVRENTSPKANKEILDEAYVMAGVSPYVSRLLIGICTSTVQLVLTQIMPYGCL 800
DB	593 KIPVAIKILNETGPRANVEFMDEALIMASMDHPHLVRLGVLCSPTIQLVTLQHPHGL 652
QY	801 LDHVRNRLGQDILLNWCQIAKMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGL 860
DB	653 LDYVHEHKDNGISOLLNWCQIAKGMVYLEERLVRDLAARNVLVKSNNHVKITDFGL 712
QY	861 ARLLDIDETEHADGGKVPKMMALESILRRRRETHQSDVMSYGVTVWELMTFCAPYDGI 920
DB	713 ARLLGDEKEYNADGGKVPKMMALESILRRRRETHQSDVMSYGVTVWELMTFCAPYDGI 772
QY	921 PAREIPDLLEKGERLPQPPCTIDVYIMVVKMIDSECRPRFELVSEFSRMRAPQRF 980
DB	773 PTEIPDLLEKGERLPQPPCTIDVYIMVVKMIDADSRPKFELAAEFMRAPQRY 832
QY	981 VVIONED-LGPASPLDSTFRSLLEDGDLVDABEYLVPOQFFCPDPAPAGAGMWHH 1039
DB	833 LVIQGDDRMKLPSNDSKFFQNLDEEDLEDMDABEYLV-PQAFNIPPIYTSRTRIDS 891
QY	1040 RHRSSTRSGGDLTLGLEPSEEAAPRS--PLAP-SEGAGSDVDFDGLGMAAGLQSLP 1096
DB	892 NRNOFYVRDGYAAEQGV-PMPYRACGIIPEAPVAGGATAEIPEDTCCNGTKURKQVATL 950
QY	1097 THDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNPQDVRPQPPSPREG 1149
DB	951 AKEDSSTQRYSDPTVPIPERVIRGELDEGDYTMPRDKPKTDYLPNVEENPFVSRKNG 1010
QY	1150 PLPAA-RPAGATLERAKTSLSPGNKVGVKDF-----AFGAVENPEYLTPOGGAAPQ 1200

DB	1011 DLQAVDNPEYHN-----APNQCPKADEYVNEPLYLNTFANTLENAEYL-----K 1055
QY	1201 PHPPAPAFSAPFONLYWQDQDPPERGA--PSPFKGTPT-----AENPEYL 1243
DB	1056 NNLPEKAKAFONPDYWNHSLPSTRLQHPDYLQEYSTKYFYKQNGRIRPIVAENPEYL 1114
RESULT 6	
ID	P79754 PRELIMINARY; PRT: 1328 AA.
AC	P79754;
DT	01-MAY-1997 (Tremblrel. 03, Created)
DT	01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Erb83;
OS	Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC	Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC	Tetraodontidae; Takifugu.
OX	NCBI_TaxID=31033;
RN	[1]_TaxID=31033;
RP	SEQUENCE FROM N.A.
RX	MEDLINE=99177347; PubMed=10077531;
RA	Gellner K., Brenner S.;
RT	"Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT	rubripes.";
RL	Genome Res. 9:251-258 (1999).
DR	EMBL: AF056116; AAC34391.1; -.
DR	HSSP: P11362; IFGK.
DR	InterPro: IPR000494; EGFR_L domain.
DR	InterPro: IPR000719; Euk_pkinase.
DR	InterPro: IPR002174; Furin-like.
DR	InterPro: IPR001245; Tyr_pkinase.
DR	Pfam: PF00757; Furin-like; 1.
DR	Pfam: PF00069; pkinase; 1.
DR	Pfam: PF01030; Recep_L domain; 2.
DR	ProDom: PD000001; Euk_pkinase; 1.
DR	SMART: SM00261; FU_3.
DR	SMART: SM00219; TyrKC; 1.
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW	ATP-binding; Transferase.
SQ	SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
Query Match 33.2%; Score 2259; DB 13; Length 1328; Best Local Similarity 40.1%; Pred. No. 1.7e-161; Matches 516; Conservative 157; Mismatches 417; Indels 196; Gaps 35;	
QY	9 WGLLLALLPP--GAASTQ----VCTGTMKLRLPASPETHLDMRLHYQCGVVOGNLEL 62
DB	4 WRLILMCVASRLRAASSQTQEAQVCPGTQNGLSSTGSEQNYNLNKORYKCEIIMGNLEI 63
QY	63 TYLPTNWSISFLQDIOEVGYVLIHQNVRQVPLQLRIVRGTLQFEDNYVALVLDNGDP 122
DB	64 TQIESNWFDFLTKTIREVTGYVLIAMHFQEIPLGQLRVIRGNSLYRERREALSVFLN--- 120
QY	123 LNNTPVTGASPCGLRELQRLSLTEILKGVLIQRPOLCYQDTILWKDIFHKQNKALTL 182
DB	121 ----YPKDG--PSGLNQLGLMNLTEILDGQVQIINNKLRYLRYGPWVYWRDII--RNNDAPIE 173
QY	183 LIDNTRSRACHPCSPMKSGRCWGESSEDCQSLTRTVCAAGGC-ARCKGPLPTDCCHEQCA 241
DB	174 IQPNERGVCH---KSC-GNYCWGPGKQDQQLTKTVCAPOCNDRCFGTSPDCCHECA 229
QY	242 AGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTACPY 301
DB	230 AGCKGLDTCFACRFLFNDSGACVPCQPTLIYKQTFQMETPNPAKYQYGSICVSCQPT 289
QY	302 NYLSTDVGSCTLVCLPHNQEVTAEDGTOR-CEKCSKPCARVCYGLGMOYIKANSKFIKIT 360
DB	290 HFV-VDGSSCVSPPPDKMEV--ERGSORQCELCSGLCPKVCBGTGAE-----QRQTVDS 342

```
QY 361 ELE-FAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFEETLEBITGYLISAWPDSL 419
Db 343 NIDSFNCTKIOGSLHFLVTGILGDDFKNVPPLDADAKKLEVFVRVREITDILNIQSPKEL 402
QY 420 PDLVSFQNIQVIRGRILHNGAYSLTLQGLGSLWGLRSLRSLGSLALIHNNTHLCFVHT 479
Db 403 NDLSVFSSLTIOGRSLFKRSLVMWRIPLTLSGLRSLREISDGSVYISQNAHLCHHT 462
QY 480 VPWDQLFRNPH-QALLHTANREDECVGEGLAGLQJLCARGHCWGPQPTQVCNCSQFLRQ 538
Db 463 VNMOTLFRGSRVRANSLSNRPMACVADGRVCDPLCSDSGCWGPGPDQCLSCRNSRHG 522
QY 539 ECVEECRLVQLPREVYNARH-CLPCHPECPONGSVTCFGEADOCVACAHYKDPFCV 597
Db 523 TVAGCHFNGSLIPREFAGLNGVCVACHPECKPOTGKASCTGPGADCEMACTFRDGPYCM 582
QY 598 ARCPGKVPDLSPYMIKPPDEGACQPCPINCTHSCVDLDKGCPEQR----ASPLTS 653
Db 583 SSCPAGVN-DGEKGLIFKFPNREHCEPCHONCTQCGSGPLNCLCEAARLTISSQITG 641
QY 654 IOY-IKAN-----SKFIGITELKROQKI-RKYMRRLLQETELVEPLTPSGAMPNQAO 705
Db 642 IALGVPAGLIFCLVLFGLM--LYHRGLAIRKRAMRYLESSESPEPLGP-GEKGTKVH 698
QY 706 MRILKTELKVKVLCGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKETLDEA 765
Db 699 ARILKPSDLRKIKPLSGVGTGKFWIPEGETVKIPVAIKTIQSSGRQTFEITDHL 758
QY 766 YMVAGVSPYVRLGICLTSTVQLVTQMPYGCLLDHRVNRGRIGSDLLNWCNQIAK 825
Db 759 LMSGLDHPYIVRLGICPGTCLQVLTQSSHGSLLHROKHTSLDPQRLNWCNQIAK 818
QY 826 GHSYLEDVLRDLAARNVLSKPNHVKITDFGLARLLDIDETEHADGGKVPKMMAL 885
Db 819 GMYLLEHVRVHKNLAARILLKNDYQVQISYGVADLLYPDKKYVSETKTPKMMAL 878
QY 886 ESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTIDV 945
Db 879 ESILFRYTHQSDVMSYGVTVWEMSGFAEPYASVQBPQVPSVLEKGERLSQPAICTIDV 938
QY 946 YMIWKMMIDSECRPRFRELVSFMRMARDPQRFVVIQNEODLGASPLDSTFYRSLLED 1005
Db 939 YMVWKMMIDENIRPTKELASDFTRMARDPPRYLVIRMEG-----ED 982
QY 1006 DMGDLVDAEEVLVPOQGFCDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEBAP 1065
Db 983 SGMGBFL-----RRGSR---GLLEADLEDEEE--- 1008
QY 1066 RSLAPSEGAGSDVFDGLGMG---AAKGLQSLPTHDRSPLO-----RYSEDPTV 1112
Db 1009 -----GLGDRFATPSLQSPSWSTSPQINSYMWVMTQURYD----- 1044
QY 1113 PLPSETDGYVAPLTCSPQ- EYVNO-----PDVRPQPPSPREGPL--PAAR 1155
Db 1045 -FAVSGGHIGYLPSPSPVDITRLQWYORSRLSSVRLTPDRSAFRSSREAELEDGQAO 1103
QY 1156 PAGATLERAKTSLPGKNGVWVDVAFAGGAVENPEYLTQGGGAOPHPPPAPSPADNLY 1215
Db 1104 CAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1216 YWDQPPPERGAPPSTFKGTPTAENPE 1241
Db 1123 --QQRKLSTASSPSSFKTWADEDE 1146

RESULT 7
Q9BIH9
AC Q9BIH9 PRELIMINARY; PRT; 1433 AA.
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
```

```
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RC SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lyceett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TK; 1.
DR SMART; SM00219; TyRK; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER.
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.3%; Score 1993.5; DB 5; Length 1433;
Best Local Similarity 32.2%; Pred. No. 2.2e-141;
Matches 466; Conservative 194; Mismatches 396; Indels 391; Gaps 37;

QY 26 CTGDMKLRRLRASPETHLDMLRHLVGGCVVGNLELTYPNLSASFLDIOIEVOGYVL 85
Db 1 CIGTNGRMSVPANREYHYKNLRDRYNTCTYVDGNLEITWIAITDLNLFQIHREVGYVL 60
QY 86 TAHNQVROVPLQRLRIVGTQLF-----EDNYALAVLDNGDPLNNTPTVTGASPGGLREL 140
Db 61 ISLYDLPOVILPRLQIIRGRTTFKLNKWEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGGVLIQRNPQLCYQDTILWKDI-FHKNNQLALTLDITNRSRACHPCSPMC 199
Db 105 ELPALRDLILGSGVGFNNVNLCHKMSINWEEILLAPQTSQOYTFNFSSEPVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSLRTRVCAGGCA--RCKGPLPTDCCHCQCAAGCTGPKHSDCLACUH 257
Db 165 EVG-CWGEAHNCORFSLKNCSPQSGRCFCGPKPRECCHLFCAGCGCTGPTQSDCLACKN 223
QY 258 FNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNYLSLTDVGSCTLVCP 317
Db 224 FYDDGVCKQCEPPMQIYNPTNYFWNPDPGKTAYGATCVRKCP-BHLLKNDGACVKKCPK 282
QY 318 HNQEVTABDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGITELEFAKCKIFGSLAF 377
Db 283 GKMPQNSE-----CVPCKGVCPKTCPEGI-----VHSDNIG-----NYKDCITIEGSL 329
QY 378 PESDGDPAISNT-----APLQPEQLQVETLEEITGYLYISAWPDSLPLDSVFNQIQ 429
Db 330 DQSDFGFOQVYTNFSGPRYIKIDPDRLEVFSTVKEITGFINQIAHPNFTLTLYFRNLE 389
QY 430 VIRGRILHNGAV-SLTLOGLGSLMGLRSLRSLGSLALIHNNTHLCFVHTVPMQOLFRN 488
Db 390 VVGRQLKENLFASVYIVKTSLSLEKSLKSVNKSIVLENSDLCFVEDIDWSEIKKS 449
```

```
QY 489 PHOALLHTANRPEDECVCEGLACHQLCARGCHGPGTQCVCNCSQFLRGQCEVCECRVLQ 548
Db 450 SDHEVMVQKRNATECHEEGMECESEQSKAGCWKGPEQCLECKNVKYKCKLDSCK--- 506
QY 549 GLPREY-VNARHCLPCHPECPONGSVTCFPGPADQCACAHYKDPFPCVARCP----- 601
Db 507 SLPRLYSVDSKTCODCHQECKD-----FCYGPNEEDNCSCMNKVDGRFCVACPTTKHAM 561
QY 602 -----SGVKPDLSPYPIWKFPD----- 618
Db 562 NGTCINCHKTCVGRGRPDTHAPDGCISCDAKIIGSDAKIERCLMKDESCPDGYYSYVL 621
QY 619 -EEG----- 621
Db 622 QEEGPKQLSGKAVCRKCHPRCKKCTGYPHEQFCQECTGYKKGQCEDECPQDFYANEE 681
QY 622 --ACQCPINCT-----HSCVDL-----KGCPAQ----- 646
Db 682 TRICLPCHQECRCGHLGDDHHCRLNKLFEQDPYDNATTTVCVSNCPASHPYKRPQEA 741
QY 647 -----RASPLTSIOYIKANSKF-----IGITELKRROQKIRKYTM 681
Db 742 KGTPYCSADSMOSGLRIEFTQVKIVGSMVALILLCVVGTAFLVFSRHKKKDAVKM 801
QY 682 RRLLOETELVEPLTPSGAMPNQMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVK 741
Db 802 TMLAGCEDSEPLRPSNVGNPLTKLRIKEAIRRGVLGNGAFGRVFKGVWMPMEGESVK 861
QY 742 IPVAIKVLRNTPSKANKEILDEAYVWAGVSPVSRLLGICLTSTVOLTQMLPYCCLL 801
Db 862 IPVAIKVLMEMSESSKEFLEAYIMASVEHPNLLKLLAVCMTSQMLLTQMLPLGCLL 921
QY 802 DHVRENRLGSLGSDLLNWCQIAGMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLA 861
Db 922 DYVRNKKDKIGSKALLNWSQIARGMAYLEERLVRDLAARNVLTVPSCVKITVEGLA 981
QY 862 RLIDIDETEHADGKVPKIMMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIP 921
Db 982 KLLDFDSDEYRAAGGKMPIKWLALCIRHRVFTSKSDVWAFGTIWEITYGARPYENVP 1041
QY 922 AREIPDLLEGERLPOPICTIDVYIMVVKWMIDSECRPRFRELVESEFMRMARDPORFV 981
Db 1042 AKDVPELIEIGHKLPPQDPCSLDVCILLSCWLDADARPTFKQLAETFAEKARDPGRYL 1101
QY 982 VIONEDLGPASPLDSTFYRSLLEDDDDMDLV----- 1012
Db 1102 MI-----PGDKFMRLPSTYNQDEKDLIRTLAPVMAAAAAAAGASNDVDPSTIA 1152
QY 1013 DABEYLVPQGGFFCPDPAPGAGGVHHRSSSTRSGGGDLTLGLEPSEBEAPRS----- 1067
Db 1153 ETDEYLQPKTRPSIMLPGSA-----VEPS-DEMPKSLRYCK 1188
QY 1068 -PLAP---SEGAGSDVFDGDLGMAKGLQSLPHTDPSLQRYSEDPTVPLPSETDGYVA 1123
Db 1189 DPLKPDDETDGHEV-----GVGGIR-----LNLPLDEDDYLM 1222
QY 1124 PLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVWVDVPAFG 1183
Db 1223 P-TCQSQ-----NQS-----TPG-----YMDLIGVPA 1243
QY 1184 AVNPEVL-----TPQGGAAPOHPPPAPSPAFNDLYWQDPPRGAPSTFKGT 1234
Db 1244 SVDNPEYLMGSTQAIAGLAQSGMG--PHTPP-----PPNTNGM 1280
QY 1235 PTAENPE 1241
Db 1281 PTHQHSQ 1287

RESULT .8
Q9UK79
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
```

```
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 6.2e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELAAACRGLLALLPPGAASCTGCTDMKLRPASPTHLDMLRHLVGGCOVQGNL 60
Db 1 MELAAACRGLLALLPPGAASCTGCTDMKLRPASPTHLDMLRHLVGGCOVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGVLYAHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLYAHNOVRQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLNNTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
Db 121 DPLNNTPTVGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYODTTLWKDIFHKNOLA 180
QY 181 LTLDITNRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKPLPTDCCHEQC 240
Db 181 LTLDITNRSRACHPCSPMKSGSCWGESSEDCOSLRTVTCAGGCARCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGL 345
Db 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSPCARVCYGL 345

RESULT 9
Q8R2X1 PRELIMINARY; PRT; 367 AA.
ID Q8R2X1
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
```

```
DR EMBL; BC027080; AAH27080.1; -.
KW Hypothetical protein.
SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match      25.6%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 4.8e-133;
Matches 323; Conservative 11; Mismatches 23; Indels 0; Gaps 0;

QY 883 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 942
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60

QY 943 IDVYIMVWKWIMIDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1002
DB 61 IDVYIMVWKWIMIDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1003 LEDDDMGDLVDAEYLVPOOGFFCDPPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEE 1062
DB 121 LEDDDMGDLVDAEYLVPOOGFFCDPPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEE 180

QY 1063 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYV 1122
DB 181 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYV 240

QY 1123 APLTCSPOPEYVNDVPRPQPSREGPLPAARPAGATLERAKTSLPGKNGVVKDVFAFG 1182
DB 241 APLTCSPOPEYVNDVPRPQPSREGPLPAARPAGATLERAKTSLPGKNGVVKDVFAFG 300

QY 1183 GAVENPEYLTPOGGAPOHPHPPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 1242
DB 301 GAVENPEYLTPOGGAPOHPHPPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 360

QY 1243 LGLDVVPV 1249
DB 361 LGLDVVPV 367

RESULT 10
Q8WYV0 PRELIMINARY; PRT; 412 AA.
AC Q8WYV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE Hypothetical 44.7 kDa protein.
GN PP3659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
RA Wan D.F., Gu J.R.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth."
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF318349; AAL55856.1; -.
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02757; YLP; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3E27D2BC CRC64;

Query Match      25.0%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 7.8e-120;

QY 569 PQNGSVTCFGPEADQCVAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
DB 1 PQNGSVTCFGPEADQCVAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
```

```
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 883 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 942
DB 1 MALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60

QY 943 IDVYIMVWKWIMIDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 1002
DB 61 IDVYIMVWKWIMIDSECRPRFRELSEFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSL 120

QY 1003 LEDDDMGDLVDAEYLVPOOGFFCDPPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEE 1062
DB 121 LEDDDMGDLVDAEYLVPOOGFFCDPPAPGAGVMVHRRSSSTRSGGDLTLGLEPSEE 180

QY 1063 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYV 1122
DB 181 EAPRSPPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPLPSETDGYV 240

QY 1123 APLTCSPOPEYVNDVPRPQPSREGPLPAARPAGATLERAKTSLPGKNGVVKDVFAFG 1182
DB 241 APLTCSPOPEYVNDVPRPQPSREGPLPAARPAGATLERAKTSLPGKNGVVKDVFAFG 300

QY 1183 GAVENPEYLTPOGGAPOHPHPPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 1242
DB 301 GAVENPEYLTPOGGAPOHPHPPAPAFDNLVYNDODPPERCAPPSTFKGTPTAENPEY 360

QY 1215 YVWD-QDPPER-----GAPPSTFKGTPTAEN 1239
DB 361 WVWTCQCEPQGVRRSPDVSSGREGLTSGAKRWEGPPTTSRGTCARN 410

RESULT 11
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian tous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnsson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities."
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1AGS.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro_M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro_M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFE1D63 CRC64;

Query Match      24.7%; Score 1677; DB 15; Length 729;
Best Local Similarity 54.0%; Pred. No. 6.5e-118;
Matches 352; Conservative 69; Mismatches 125; Indels 106; Gaps 14;

QY 569 PQNGSVTCFGPEADQCVAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
DB 1 PQNGSVTCFGPEADQCVAHYKDPFVCVARCPGVKPDLSYMPIWKFLPESGACQPCPI 628
```



Db 141 PEETATPKTGP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHP 197  
 QY 629 NCHTSCVDLDDKCPAQRASPLTSIQYIKANSKF-----IGITELKRROOKIRKYTW 682  
 Db 198 NTRCGKGLGECPP---NGSKTPTSIAAGVVGGLCLLVVGGIGLYLRRHIVKRTLR 254  
 QY 683 RLLOETELVEPTTPSGAMPNQAOMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKI 742  
 Db 255 RLLOERELVEPTTPSGEAPNQAHLRIKETEFKVKVLGSGAGFTVYKGLWIPGEKVKI 314  
 QY 743 PVAIKVLRENTSPKANKEILDEAYMAGVGSPPVSRLLGICLTSTVQLTQMLPYGCLLD 802  
 Db 315 PVAIKELRENTSPKANKEILDEAYMAGVGSPPVSRLLGICLTSTVQLTQMLPYGCLLD 374  
 QY 803 HVYENRGRIGSDDLNMCMQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLAR 862  
 Db 375 YIREHKNDIGSYQLLWNCVQIAGMNYLEERRLVHRDLAARNVLKSPNHVKITDFGLAR 434  
 QY 863 LLDIDETEHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPA 922  
 Db 435 LLGADEKEYHAEGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGSKPYDGIPA 494  
 QY 923 REIPDLLEKGERLPQPICTIDVYIMVVKCWMIDSECRPRFRELVSFGRMARDPQRFV 982  
 Db 495 SEISSVLEKGERLPQPICTIDVYIMVVKCWMIDSECRPRFRELVSFGRMARDPQRFV 554  
 QY 983 IO-NEDLGASPLDSTFYRSLLDDMDGLVDAEEYLVPOQGFCCPDAPAGAGMVYHRH 1041  
 Db 555 IOGERHMLPSPDTSKFRYTLMEEDMEDIVDAEYLVPHQGF-----598  
 QY 1042 RSSTSRGGGDLTLGLEPSEEAAPRSP-----APSEGAGSDVFDGLGMAAKGQSLP 1096  
 Db 599 NSPST-----SRTPLLSLSATSNNATNCID-----RNGQGP 632  
 QY 1097 THDPSPLORYSEDPVLPSET--DGYVAPLTCSPQPEYVQNDVDPQPPSPREGPLPAA 1154  
 Db 633 VREDSFVQRYSSDPTGNFLEESIDDGFL-----PAPEYVQ--LMPKKPS-----675  
 QY 1155 RPAGATLERAKTSPCKNGVVKDVF-----AFGAVENPEYL 1191  
 Db 676 -----TAMVQNIYNNISLTATSKLPMDSRYQNSHSTAVDNPEYL 715  
 RESULT 12  
 Q86714  
 ID Q86714 PRELIMINARY; PRT; 567 AA.  
 AC Q86714;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE V-erbB protein (Fragment).  
 GN V-ERBB.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.  
 OX NCBI\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities.";  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL; S69372; AAC60727.1; --  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Tyrosine-protein kinase.  
 FT NON\_TER 1  
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;  
 Query Match 24.6%; Score 1675; DB 15; Length 567;  
 Best Local Similarity 54.6%; Pred. No. 6.3e-118;  
 Matches 351; Conservative 67; Mismatch 119; Indels 106; Gaps 14;  
 QY 578 GPEAQCACVACAIYKDPFPCVACRPSGVKPDLSYMPFIWKPFDEGACQPCINCTHSCVDL 637  
 Db 1 GP--DHCMKCAHFIDGPHCVKACPAVLGENDTL-VMKYADANAVCOLCHPNTCKGK 57  
 QY 638 DDKGCPAEQASPLTSIQYIKANSKF-----IGITELKRROOKIRKYTWRRLLQSTELV 691  
 Db 58 GUEGCP---NGSKTPTSIAAGVVGGLCLLVVGGIGLYLRRHIVKRTLRRLLORELV 114  
 QY 692 EPLTPSGAMPNQAOMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLR 751  
 Db 115 EPLTPSGEAPNQAHLRIKETEFKVKVLGSGAGFTVYKGLWIPGEKVKIPVAIKELRE 174  
 QY 752 NTPSKANKEILDEAYMAGVGSPPVSRLLGICLTSTVQLTQMLPYGCLLDHVRENRGL 811  
 Db 175 ATSPKANKEILDEAYMAGVGSPPVSRLLGICLTSTVQLTQMLPYGCLLDVIREHKNI 234  
 QY 812 GSODLLNMCQIAKMSYLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEV 871  
 Db 235 GSQYLLNMCVQIAGMNYLEERRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEV 294  
 QY 872 HADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEK 931  
 Db 295 HAEGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGSKPYDGIPASEISSVLEK 354  
 QY 932 GERLPQPICTIDVYIMVVKCWMIDSECRPRFRELVSFGRMARDPQRFVIO-NEDLGP 990  
 Db 355 GERLPQPICTIDVYIMVVKCWMIDSECRPRFRELVSFGRMARDPQRFVIO-NEDLGP 414  
 QY 991 ASPLDSTFYRSLLDDMDGLVDAEYLVPOQGFCCPDAPAGAGMVYHRHSSSTSGG 1050  
 Db 415 PPTOSKFRYTLMEEDMEDIVDAEYLVPHQGF-----NSPST-----454  
 QY 1051 GDLTLGLEPSEEAAPRSP-----APSEGAGSDVFDGLGMAAKGQSLPLTHDPSPLOR 1105  
 Db 455 -----SRTPLLSLSATSNNATNCID-----RNGQHPVREDSFVQR 492  
 QY 1106 YSEDPVLPSET--DGYVAPLTCSPQPEYVQNDVDPQPPSPREGPLPAARPAGATLER 1163  
 Db 493 YSSDPTGNFLEESIDDGFL-----PAPEYVQ--LMPKKPS-----526  
 QY 1164 AKTLPCKNGVVKDVF-----AFGAVENPEYL 1191  
 Db 527 -----TAMVQNIYNNISLTATSKLPMDSRYQNSHSTAVDNPEYL 566  
 RESULT 13  
 Q64895  
 ID Q64895 PRELIMINARY; PRT; 962 AA.  
 AC Q64895;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Gag v-erb-A, v-erb-B protein.  
 GN GAG, V-ERB-A, V-ERB-B.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.  
 OX NCBI\_TaxID=11861;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90206603; PubMed=1969616;  
 RA Bruskina A., Jackson J., Bishop J.M., McCarley D.J., Schatzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 RT transforming potential of the oncogene v-erb-B.";

CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC	-1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR	EMBL; X52209; CAA36459.1; -
DR	EMBL; X52211; CAA36459.1; JOINED.
DR	HSSP; P10828; 2NLL.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR000536; Hormone_rec_lig.
DR	InterPro; IPR001723; Stdhmn_receptor.
DR	InterPro; IPR001245; Tyr_kinase.
DR	InterPro; IPR001628; Znf_C4steroid.
DR	Pfam; PF00104; hormone_rec; 1.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF00105; zf-C4; 1.
DR	PRINTS; PR00398; STRDHOMONER.
DR	PRINTS; PR00047; STROIDFINGER.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	ProDom; PD000035; Znf_C4steroid; 1.
DR	SMART; SM00430; HOLI; 1.
DR	SMART; SM00219; TyrKc; 1.
DR	SMART; SM00399; ZnF_C4; 1.
DR	PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW	ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW	Transcription regulation; Transferase; Tyrosine-protein kinase;
KW	Zinc-finger.
SQ	SEQUENCE 962 AA; 108320 MW; 3C5AED791BAE95CE CRC64;
Query Match            23.7%; Score 1610.5; DB 15; Length 962;	
Best Local Similarity 50.9%; Pred. No. 1e-112;	
Matches 352; Conservative 67; Mismatches 149; Indels 123; Gaps 17;	
Qy	541 VEECRVQLGPRE-VYNAR-HCLP-----CHPEQC 568
Dd	541 IEKQESYLAFEHYINRYKHNIHFHWSKULMKVADLRMI GAYHASRFLHMKECPTELS 413
Qy	569 PQNGSVTCFGEAPDQCACAHYKDPPFCVCARCPGVKPDLSYMPIWKFPDEEGAGOCPCPI 628
Dd	414 PQE-----VGP--DHCMKCAHFIDGPHCVKACPAGVLGENDTL-VVKYADANAVCQLCHP 465
Qy	629 NCTHSCVDLDDKGCPAEQRASPLTSIQYIKANSKF-----IGITELKRQQKIRKYVTMR 682
Dd	466 NCTRGCCKGPLEGCP---NGSKTPSIAGVVGGLLCLLVVGLGIGLYLRHRHVIRKRTLR 522
Qy	683 RLLOETELVEPLTSPGAMPNOAQNRILKETELRKVKVLGSGAGTVYKGIWIPDGENVKI 742
Dd	523 RLLOERELVEPLTSPGAPNOAHLRIKETEFKKVKVLGFAGFTYKGLWIPEGEKVTI 582
Qy	743 PVAIKVLRENTSPKANKEIIDEA VMAGVGSYPVSRLLGICLTSTVOLVTOLMPYCGCLLD 802
Dd	583 PVAIKELREATSPKANKEIIDEA VMASVDNPHVCRLLGICLTSTVOLITQLMPLYCGCLLD 642
Qy	803 HVRENRLGSGDLLNNCMQIAKMSYLEVDRLVHRDLAARNLVKSPNHVKITDFGLAR 862
Dd	643 YIREHKDNISQYLLANCVCQIAKGMNYLEERHVRDLAARNLVKTPOHVKITDFGLAK 702
Qy	863 LLDIDETEXHADGKVPICKMALESILRRFRTHQSODVMSYGVTVMELMTFGAKPYDGIPA 922
Dd	703 QLGADEKEYHAEGKVPICKMALESILHRYTHQSODVMSYGVTVMELMTFGSKPYDGIPA 762
Qy	923 REIPDLLEKERLPQPPICTIDVYMIMVKCMMDSECRPFRELVSFEFSRMARDPQRFVU 982
Dd	763 SEISSVLEKERLPQPPICTIDVYMIMVKCMSGASRPKRELIIAEFSKNARDPPRYLV 822
Qy	983 IQ-NEDLGASPILDSTFYRSLLDDMGDLVDAAEYLVFQQGFFCPDPAPGAGGVVHHRH 1041
Dd	823 IOGDERMHLPSPTDSKPYRTLMEEDMEDIVDAEYLVHQGF-----NSATKICDRNGGH--- 898
Qy	1042 RSSSTSRCGGDLTLGLEPSEEAPRPLAPSEGAGSDVFDGLGMAAAGLOSPLPHDPS 1101
Dd	867 NSPST-----SRTPLLSSLSATS-----NSATKICDRNGGH--- 898

```

Qy 991 ASPLDSTFRSLLEDDMDGLVDAREYLVPOQGFPCPDAPGACGWMVHRHSSSTRSG 1050
Db 415 PSPTDUSKFRITLMESEDMEDVDAEYLVPHGFF-----NSPT----- 454
Qy 1051 GDLTLGLFSEBEAPRSP-----APSEGAGSDVFDGDLGMAAGKLSLPHDPSPLO 1105
Db 455 -----SRTPLLSLSATSNSNATNCIDRNG-----H----- 481
Qy 1106 YSEDPTVPLPSETDGVAPLTCSPROPEYVNDVFPQPSREGPLPAARPAAGAT-LERA 1164
Db 482 -----PVREDGFL-----PAPEYVNO--LMPKXPSTAMVOIQIYNISLTAISK 524
Qy 1165 KTLSPGKNGVWKFVAFGAVENPEYL 1191
Db 525 PMSRYQN-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5
ID Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1]_
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Mailhe N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
[2]
SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7TAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mailhe N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Niehi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saio T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saio R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Sakimi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;

```

```

RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGI; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ
SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match      22.2%; Score 1506.5; DB 11; Length 655;
Beat Local Similarity 44.5%; Pred. No. 4.2e-105;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

Qy 11 LLLALLPPGAA--STQVCTGTDMLRLPASPTHLDMLRHLYQGCVQVGNLELYLPTN 68
Db 14 LLTALCAAGGAALEKKVCQGTNSRLTQLGTFEDHFLSLQRMYNCEVVLGNLEITYVQRN 73

Qy 69 ASLSFLDIOIEVQGYLIAHNOVROVPLQRLRIVRGTOLFEDNYALAVLDNGDPLNNTTP 128
Db 74 YDLSEKTIQEVAGYVLIANTVERIPLENQIIRGNALYENTYALAILSN----- 124

Qy 129 VTGASPGGLRELQLASLTILKGGVLIQRNPOLCYQDTILWKDI----FHKNNQLALTIL 184
Db 125 -YGNRTGLRELPMNLQELIGAVFSNNPILCNMDTIQWRDIQNVFMSMDL---- 180

Qy 185 DTNRSRACHPCSPMCKSGRSGESSEDCOSLTRTVACGCA-RCXGPLPTDCHEGCAAG 243
Db 181 -QSHPSSCPCKDPCSPNGSCWGGEGENCQKLTIIICAQCSHRCGRSPSCCHNQCAAG 239

Qy 244 CTGPKHSDCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRVTFGASCVTACPNY 303
Db 240 CTGPRESDCLVCKQFQDEATCKDTCTPPLMLYNTPTTYQMDVNPGEKYSFGATCKVKCPRY 299

Qy 304 LSTDVGSCTLVCLHNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGITELE 363
Db 300 VTDHSGSVACRACPDYEV-BEDGIRKCKCDGPKCKVCGNGIGIGEFK-DTLSINATNIK 357

Qy 364 -FAGCKKIFGSLAFLPESFDGDPDANTAPLQPEQLQVFTLEETIGLYLISAWPDSLPL 422
Db 358 HFKYCTAISGDHLILPVAFKGDSFTRTPPLDPRELEILKTKVEITGFLLIQAWPDNWTDL 417

Qy 423 SVFQNLQVIRGILHNGAYSILTLOGIGISWGLRSLRELGLALIHNTHLCHFVHTVPW 482
Db 418 HAFENLEIRGTRTKHQGFSLAVVGNLITSLGRSLKESIDGVDVIISGRNRLCYANTINW 477

Qy 483 DOLFENPHQALILHTANRDECEVCEGLACHQLCARGHCKMGPGPTQCVNCSQFLRGCEC 542
Db 478 KKLFTGTPNQTKIMNRAEKCKAVNHVCNPLCSSEGCWGPEDPCVSCQNVSRGECV 537

Qy 543 ECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFCVARCPS 602
Db 538 KCNILEGEPREFENSECTQCHPECLPQAMNITCTGRGPDNCIQCAHYIDGPHCVKTC 597

Qy 603 GVKPDLSYMTPKFDEECACQPCPINCTHSCVLDLDDKGC 642
Db 598 GIMGENNTIL-VWKYADANNVCHLANCTYGCAGPGLQGC 636

```

Search completed: July 22, 2003, 09:24:14  
Job time : 51.6008 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.5911 Seconds  
(without alignments)  
4891.279 Million cell updates/sec

Title: SEQ4-632-652-12

Perfect score: 6796

Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYGLDVPV 1249

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6619	97.4	1255	1	P04626 homo sapien
2	5857	86.2	1257	1	P06494 rattus norv
3	5833.5	85.8	1254	1	P00533 mesocricetu
4	3113	45.8	1210	1	P00533 homo sapien
5	3083	45.4	1210	1	Q01279 mus musculu
6	2925.5	43.0	1308	1	Q15303 homo sapien
7	2907	42.8	1308	1	Q62956 rattus norv
8	2651.5	39.0	1167	1	P13388 xiphophorus
9	2407.5	35.4	1342	1	P21860 homo sapien
10	2320.5	34.1	1339	1	Q62799 rattus norv
11	1944	28.6	1426	1	P04412 drosophila
12	1706.5	25.1	634	1	P00534 avian leuko
13	1660	24.4	604	1	P00535 avian eryth
14	1587	23.4	540	1	P11273 avian eryth
15	1552	22.8	703	1	P13387 gallus gall
16	1282	18.9	1323	1	P24348 caenorhabdi
17	1142.5	16.8	245	1	P70424 mus musculu
18	711	10.5	1363	1	Q02466 branchiost
19	690	10.2	1300	1	Q02414 mus musculu
20	679	10.0	1382	1	P06213 homo sapien
21	679	10.0	1607	1	Q25410 lymnaea sta
22	678.5	10.0	1300	1	P14617 cavia porce
23	678	10.0	1297	1	P14616 homo sapien
24	678	10.0	1383	1	P15127 rattus norv
25	677.5	10.0	1372	1	P15208 mus musculu
26	673	9.9	1477	1	Q25197 hydra atten
27	632	9.3	1367	1	P08069 homo sapien
28	618	9.1	1373	1	Q60751 mus musculu
29	615	9.0	1390	1	Q93105 aedes aegypt
30	614.5	9.0	1370	1	P24062 rattus norv
31	612	8.7	2146	1	P09208 drosophila
32	588	8.0	757	1	P53356 hydra atten
33	581.5	8.6	1053	1	Q00944 gallus gall

Q07494 gallus gall  
P07949 homo sapien  
P09759 rattus norv  
Q91738 xenopus lae  
P54762 homo sapien  
P54760 homo sapien  
P34152 mus musculu  
Q35346 rattus norv  
Q91736 xenopus lae  
P00521 abelson mur  
Q05397 homo sapien  
Q03145 mus musculu

34 576.5 8.5 984 1 EPB1\_CHICK  
35 576 8.5 1114 1 RET\_HUMAN  
36 574.5 8.5 984 1 EPB1\_RAT  
37 569 8.4 1068 1 FAK1\_XENLA  
38 568.5 8.4 984 1 EPB1\_HUMAN  
39 568 8.4 987 1 EPB4\_HUMAN  
40 568 8.4 1052 1 FAK1\_MOUSE  
41 568 8.4 1055 1 FAK1\_RAT  
42 567 8.3 902 1 EPBB\_XENLA  
43 565.5 8.3 746 1 ABL\_MLVAB  
44 565 8.3 1052 1 FAK1\_HUMAN  
45 557.5 8.2 977 1 EPA2\_MOUSE

ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 13-AUG-1987 (Rel. 05, Last annotation update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
DE surface receptor HER2) (MLN 19).  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, C-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
CC POTENTIAL NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
CC ALPHA AND AMPHIREGULIN.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC -!- (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC CC RESIDUES (BY SIMILARITY).  
 CC CC -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN  
 CC CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC CC -----  
 CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC CC the European Bioinformatics Institute. There are no restrictions on its  
 CC CC use by non-profit institutions as long as its content is in no way  
 CC CC modified and this statement is not removed. Usage by and for commercial  
 CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC CC -----  
 CC CC EMBL; M11767; AAA35808.1; JOINED.  
 CC CC EMBL; M11761; AAA35808.1; JOINED.  
 CC CC EMBL; M11762; AAA35808.1; JOINED.  
 CC CC EMBL; M11763; AAA35808.1; JOINED.  
 CC CC EMBL; M11764; AAA35808.1; JOINED.  
 CC CC EMBL; M11765; AAA35808.1; JOINED.  
 CC CC EMBL; M11766; AAA35808.1; JOINED.  
 CC CC EMBL; M11730; AAA75493.1; -.  
 CC CC EMBL; M12036; AAA35978.1; -.  
 CC CC EMBL; X03363; CRA27060.1; -.  
 CC CC PIR; A25491; A25491.  
 CC CC PIR; A24571; A24571.  
 CC CC HSSP; P11362; LFCK.  
 CC CC Genew; HGNC:3430; ERBB2.  
 CC CC MIM; 164870; -.  
 CC CC InterPro; IPR000494; EGFR\_L\_domain.  
 CC CC InterPro; IPR000719; Euk\_pkinase.  
 CC CC InterPro; IPR002174; Furin-like.  
 CC CC InterPro; IPR001245; Tyr\_pkinase.  
 CC CC InterPro; IPR004019; YLP\_motif.  
 CC CC Pfam; PF00069; pkinase; 1.  
 CC CC Pfam; PF00757; Furin-like; 1.  
 CC CC Pfam; PF01030; Recep\_L\_domain; 2.  
 CC CC Pfam; PF02757; YLP; 2.  
 CC CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC CC SMART; SM00261; FU; 3.  
 CC CC SMART; SM00219; TyrKc; 1.  
 CC CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 CC CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 CC CC PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
 CC CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC CC Polymorphism.  
 CC CC SIGNAL 1 21  
 CC CC CHAIN 22 1255  
 CC CC DOMAIN 22 652  
 CC CC TRANSMEM 653 675  
 CC CC DOMAIN 676 1255  
 CC CC DOMAIN 720 987  
 CC CC NP\_BIND 726 734  
 CC CC BINDING 753 753  
 CC CC ACT\_SITE 845 845  
 CC CC DISULFID 195 204  
 CC CC DISULFID 199 212  
 CC CC DISULFID 220 227  
 CC CC DISULFID 224 235  
 CC CC DISULFID 236 244  
 CC CC DISULFID 240 252  
 CC CC DISULFID 255 264  
 CC CC DISULFID 268 295  
 CC CC DISULFID 299 311  
 CC CC DISULFID 315 331  
 CC CC DISULFID 334 338  
 CC CC D-SULFID 511 520  
 CC CC DISULFID 515 528  
 CC CC DISULFID 531 540  
 CC CC DISULFID 544 560

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	634	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTid=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	/FTid=VAR_004078.
FT	CONFLICT	1255	1255	P -> A (IN REF. 2).
SQ	SEQUENCE	1255	AA; 137909	MM; 39E9DFDA04DCF962 CRC64;

Query Match 97.4%; Score 6619; DB 1; Length 1255;  
 Best Local Similarity 96.9%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 17; Indels 16; Gaps 2;

Qy	1	MELALCRWGLLLALLPPGAAS	TOVCTG	TDMLRLPAS	PETHLDMRLHLYG	CCVQGNL	60
Db	1	MELALCRWGLLLALLPPGAAS <td>TOVCTG</td> <td>TDMLRLPAS <td>PETHLDMRLHLYG <td>CCVQGNL</td> <td>60</td> </td></td>	TOVCTG	TDMLRLPAS <td>PETHLDMRLHLYG <td>CCVQGNL</td> <td>60</td> </td>	PETHLDMRLHLYG <td>CCVQGNL</td> <td>60</td>	CCVQGNL	60
Qy	61	ELTYLPTNASLSFLQD	IQEVGYVLI	AHNOVRQV	PLQRLRIVRG	TQLPEDNYALAVL	120
Db	61	ELTYLPTNASLSFLQD	IQEVGYVLI	AHNOVRQV	PLQRLRIVRG	TQLPEDNYALAVL	120
Qy	121	DPLNNTTPVTGAS	PGGLRELQ	RLSLTEILKGGVLI	IQNPQLCYOD	TLWKDIFHRNNQ	180
Db	121	DPLNNTTPVTGAS	PGGLRELQ	RLSLTEILKGGVLI	IQNPQLCYOD	TLWKDIFHRNNQ	180
Qy	181	LTLIDTNRSRACH	PCSPKCGSR	CGWSSDQSL	TRTVCA	GAGCARCKG	240
Db	181	LTLIDTNRSRACH	PCSPKCGSR	CGWSSDQSL	TRTVCA	GAGCARCKG	240
Qy	241	AAGCTGPKGSDCL	ACHFNHSGI	CELHCPAL	VTYNTDT	FESMPNPEGR	300
Db	241	AAGCTGPKGSDCL	ACHFNHSGI	CELHCPAL	VTYNTDT	FESMPNPEGR	300
Qy	301	YNYLSTDVGS	CTLVCP	LNHNEVTA	EDGTQRC	EKCKPCAR	360
Db	301	YNYLSTDVGS	CTLVCP	LNHNEVTA	EDGTQRC	EKCKPCAR	360
Qy	361	ELFAGCKIFGSL	AFIPES	FDGDPAS	NTAPLQ	PEQLOVFT	420
Db	361	ELFAGCKIFGSL	AFIPES	FDGDPAS	NTAPLQ	PEQLOVFT	420
Qy	421	DLVSFQNLQVIR	GLHNGAY	SLTQGLG	ISWGLRSL	RELSGSLALH	480
Db	421	DLVSFQNLQVIR	GLHNGAY	SLTQGLG	ISWGLRSL	RELSGSLALH	480
Qy	481	PWDQFRNP	PHOALLHTAN	RRPEDE	CVGEG	LACHQI	540
Db	481	PWDQFRNP	PHOALLHTAN	RRPEDE	CVGEG	LACHQI	540
Qy	541	VEECRVLQGL	PREVYNAR	HCLPCH	PECPONG	SVTCFGE	600
Db	541	VEECRVLQGL	PREVYNAR	HCLPCH	PECPONG	SVTCFGE	600
Qy	601	PSGVKPDLSY	MPIMKFP	DEEGAC	QPCP	INCTHSCVD	660
Db	601	PSGVKPDLSY	MPIMKFP	DEEGAC	QPCP	INCTHSCVD	660
Qy	661	SKFTGI	-----	TELKRR	QOKIRKY	TMRLLO	709
Db	661	SKFTGI	-----	TELKRR	QOKIRKY	TMRLLO	709
Qy	709	SAVVGILLV	VVGLV	FGILIK	RRQOKIRKY	TMRLLO	715
Db	709	SAVVGILLV	VVGLV	FGILIK	RRQOKIRKY	TMRLLO	715



```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 86.2%; Score 5857; DB 1; Length 1257;
Best Local Similarity 85.8%; Pred. No. 2.6e-301;
Matches 1083; Conservative 51; Mismatches 110; Indels 18; Gaps 4;

QY 1 MELAAALCRWGLLALLPPGAASCTOVCTGTOMKRLRLPASPTHLDMLRHLRYOGQVQGNL 60
DB 1 MELAAWCRWGLLALLPPGAGTQVCTGTOMKRLRLPASPTHLDMLRHLRYOGQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIQEVQGYVLIAHNQVQVPLQRLRIVRGTLQFEDNYALAVLDNG 120
DB 61 ELTYVVPANASLSFLQDIQEVQGYVLIAHNQVQVPLQRLRIVRGTLQFEDKYALAVLDNR 120
QY 121 DPLNNTPTVT-GASPGGLRELOLRSLTEILKGGVLIQORNPOLCYQDTILKWDIFHKKNQL 179
DB 121 DPQDNVAASPTPGTREGLEQLRLSLTEILKGGVLIQIRGNPOLCYQDMVLWKVDVFRKNQL 180
QY 180 ALTLIDITNRGRACHPCSPCKMGRCWGESSEDCQSLRTVTCAGGCARCKPLPTDCCHEQ 239
DB 181 APVIDITNRGRACHPCAPACKDNHCWGESSEDCQILGTICTSCCARCKRLPTDCCHEQ 240
QY 240 CAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMPNPGRYTFFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACILHFNHSGICELHCPALVTYNTDTFESMHNPGRYTFFGASCVTTC 300
QY 300 PYNYLSTDVSGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCYGLQWQVYKANSKEFIGI 359
DB 301 PYNYLSTEVSGSCTLVCPNPNQEVTAEDGTORCEKSPCARVCYGLQWHLRGARAITSD 360
QY 360 TELEFACCKIFGSLAFLPESFDGDPASNTAPLOEQVFTLEETLGYLYTSAPDLSL 419
DB 361 NVQEFDCCKIFGSLAFLPESFDGDPSSGATLRLPEQLQVFTLEETLGYLYTSAPDLSL 420
QY 420 PDLVSFQNLQVIRGRILHNGAYSITLQGLGISWGLSLRELGSGLALIHNNTHLCFVHT 479
DB 421 RDLVSFQNLRIIRGRILHNGAYSITLQGLGHSLSLRELGSGLALIHNNHLCFVHT 480
QY 480 VPMDOLEFRNPHOALLHTANPEDE-CYVEGLACHQLCARGHCWCPGTQCVNCSQFLRGQ 538
DB 481 VPMDOLEFRNPHOALLHSGNRPEDLCVSSGLVNSLCAHGCWCPGTQCVNCSHFURGQ 540
QY 539 ECVEECRVLOGLPREYVYVNRHCLPCHPECPQNGSVTCFGEADQVACAHYKDPPECVA 598
DB 541 ECVEECRVWGLPREYVYVSDKRLPCHPECPQNSSETCFGEADQVACAHYKDXSSCVA 600
QY 599 RCPGSGVKPDLSYMPIWKPDBEAGACQPCPNCTHSCVDLDDKGCAPASQASPLTSIYIK 658
DB 601 RCPGSGVKPDLSYMPIWKPDBEAGICQPCPNCTHSCVDLDERGCPAQASQASPVTFI 656
QY 659 ANSKFIGI-----TELKERQOKIRKTYMRRLLQETELVELPTBSCAMPNQOMR 707
DB 657 -IATVGVGLFLVLVVVVGILIKRRQOKIRKTYMRRLLQETELVELPTBSCAMPNQOMR 715
QY 708 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDYAV 767
DB 716 ILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDYAV 775
QY 768 MAGVGSPPVSRLLIGICLTSTVQLTQMPYGLLDHVRNRRIGLSQDOLLNWCNQIAKGM 827
```

```
DB 776 MAGVGSPPVSRLLIGICLTSTVQLTQMPYGLLDHVRNRRIGLSQDOLLNWCNQIAKGM 835
QY 828 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 887
DB 836 SYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALES 895
QY 888 ILARRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYM 947
DB 896 ILARRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVYM 955
QY 948 IMVKCMWIDSECRPRFRELVSFSESRMARDPQRVFIQNEIDLGPASPLDSTFYRSLLEDDDD 1007
DB 956 IMVKCMWIDSECRPRFRELVSFSESRMARDPQRVFIQNEIDLGPSSPMDSTFYRSLLEDDDD 1015
QY 1008 MGLDVAEYLVPOQGFCCPDPAAGGGMVHRRSRSSSTRSGGDLTLGLLEPSEEEAPRS 1067
DB 1016 MGLDVAEYLVPOQGFCCPDPAAGGGMVHRRSRSSSTRSGGDLTLGLLEPSEEGPPRS 1075
QY 1068 PLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTC 1127
DB 1076 PLAPSEGAGSDVFDGDLAMGVTKGLQSLSPHDSPLQRYSEDPTLPLPETDGYVAPLAC 1135
QY 1128 SPOPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVEN 1187
DB 1136 SPOPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKOVFAFGGAVEN 1195
QY 1188 PEYLTQCGGAAPHPHPPAPFAFPAFDNLYWDQDPPERCAPPSTFKGTPTAENPEYLGVDV 1247
DB 1196 PEYLVPRGTPASPFPSPAFPAFDNLYWDQDSSSQGPPSPNFGPTPTAENPEYLGVDV 1255
QY 1248 PV 1249
DB 1256 PV 1257

RESULT 3
ERB2_MESAU
ID_ERB2_MESAU STANDARD; PRT; 1254 AA.
AC_Q60553;
LT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Iehikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))

CC -----  
 DR EMBL; D16295; BAA03801.1; -  
 DR HSP; P11362; IFGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR004019; YLP motif.  
 DR Pfam; PF00069; pkinase.1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3  
 DR SMART; SM00219; Ty-Kc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 158 368 CYS-RICH.  
 FT DOMAIN 472 644 CYS-RICH.  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.  
 FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).  
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.8%; Score 5833.5; DB 1; Length 1254;  
 Best Local Similarity 85.3%; Pred. No. 4.4e-300;  
 Matches 1075; Conservative 60; Mismatches 108; Indels 17; Gaps 3;

Qy 1 MELAALCRWGLLLALLPPGAASCTGCTGDMKRLPASPETHLDMRLHLYOGCQVQGNL 60  
 Db 1 MELAAWCGWGLLLALLSPGASGCTGCTGDMKRLPASPETHLDMRLHLYOGCQVQGNL 60  
 Qy 61 ELTYLPTNASLSFLQDIOEQVGYVLI AHNQVRVPLQRLRIVRGTLQFEDNYALAVLNG 120  
 Db 61 ELTYLPANATLSFLQDIOEQVGYMLI AHQVRHVPLQRLRIVRGTLQFEDNYALAVLNR 120  
 Qy 121 DPLNNTPTVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYODTILWKDIFHKNNOLA 180  
 Db 121 DPLDNVTTATGRTPEGLRELQLRSLTEILKGGVLIQRNPOLCYODTILWKDIFVRKNNOLA 180  
 Qy 181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACGGCARCKGPLTDCCHQOC 240  
 Db 181 PVDIDTNRSRACPCAPACKDNHCWASPEDCQTLGTIAPRAVPAARALPTDCCHQOC 240  
 Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP 300  
 Qy 301 YNYLSTDVGSCTLVCPLNHQNVEAEDGTORCEKCKSPCARVCYGLGMOYIKANSKFIGIT 360  
 Db 301 YNYLSTVGSCTLVCPLNHQNVEAEDGTORCEKCKSKCARVCYGLGMEHLRGARITSAN 360  
 Qy 361 ELEFAGCKKIFGSLAFILPESFDGDPASNTAPLQPEQLQVFETLEITCYLYISAMPDSL 420  
 Db 361 IQEFAGCKKIFGSLAFILPESFDGNPSSGIALPTPEQLQVFETLEITCYLYISAMPDSL 420  
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLOGLIGISWGLRLSRLBSGLGLALIHNNTHLCFVHTV 480  
 Db 421 DLSVFQNLRVIRGRVLDHGYSLALQGLIRWGLRLSRLBSGLGLVLIHRNTHLCFVHTV 480  
 Qy 481 PWDOLFNPFOALLHTANRPEDECVEGELACHQICARHCWGCGPGTQCVNCSQFIRGQBC 540  
 Db 481 PWDOLFNPFOALLHSGNPSBEEGCKLDFACYLCAHGHGCGPGTQCVNCSHFIRGQBC 540  
 Qy 541 VEECRVLQGLPREVYNARHCLPCHPEQOPNGSVTCFGEADQCVACAHYKDPPECVAPC 600  
 Db 541 VKECRVMKGLPREYNGKHCLPCHPEQOPNSTETCTGSEADQCTACPHYKDSPPFCVAC 600  
 Qy 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPASPLTSIQYIKAN 660  
 Db 601 PSGVKPDLSTYMPIWKYPDEEGACQPCINCTHSCVDLDERGCPAQASPLTSI----I 655  
 Qy 661 SKFTGI-----TELKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRIL 709  
 Db 656 ATVVGILLFLVIGVVVVGILIKRRQKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRIL 715  
 Qy 710 KETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVNA 769  
 Db 716 KETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVNA 775  
 Qy 770 GVGSPPYSRLILGICLTSTVQLVTOLMPYGCILLDHVRENRLGRLGSDLLNWCQIAGMSY 829  
 Db 776 GLGSPYSRLILGICLTSTVQLVTOLMPYGCILLDHVREHRLGRLGSDLLNWCQIAGMSY 835  
 Qy 830 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETETHADGGKVPYIKWMALESTL 889  
 Db 836 LEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETETHADGGKVPYIKWMALESTL 895  
 Qy 890 RRRRTHQSDVWSYGVYVWELMTFGAKPYDGI PAREIPDLEKGERLPPOPICTIDVYIM 949  
 Db 896 RRRRTHQSDVWSYGVYVWELMTFGAKPYDGI PAREIPDLEKGERLPPOPICTIDVYIM 955  
 Qy 950 VKCMWIDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLEDDMG 1009  
 Db 956 VKCMWIDSECRPRFRELVESEFSRMARDPQRFVVIQNEDLGPSSPLDSTFYRSLLEDDMG 1015  
 Qy 1010 DLVDAEYLVFQQGFFCPDPAPAGWVHRRHSSTSGGGDLTLGLEPSEEEAPRPL 1069  
 Db 1016 DLVDAEYLVFQQGFFCPDPAPAGWVHRRHSSTSGGGELTLGLEPSEEEAPRPL 1075



```
Oy 1070 APSEGAGSDVFDGLGMAAGLQSLPHDPSPLQRYSEDTVPPLPSETDGYVAPLTCSF 1129
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1076 APSEGAGSDVEGELGMAATGKPOSISPRDLSPLQRYSEDTPLPFTETDGYVAPLACS 1135
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 1130 QPEYVQNDVVRPQPPSPREGPLPAARAGATLERAKTLSPKNGVVDVFAFGAVENPE 1189
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1136 QPEYVQNDVVRPQPLTPEGPLPVRPAGATLERPKTLSPKNGVVDVFTFGAVENPE 1195
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Oy 1190 YLTGCGNAPOPHPHPPAFSDNLYWDQDPPERGAPPSTFKGTPAENPEYLGLDVVP 1249
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1196 YLVRGGSASQPH-PPALCPAFDNLNLYWDQDPPSERGSPNTFEGTPTAENPEYLGLDVVP 1254
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 4
EGFR_HUMAN
ID EGFR_HUMAN STANDARD; PRT; 1210 AA.
AC P05333; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; O00732;
AC O06688; Q9B252; Q9B2C9; Q9GZX1; Q9H3C9;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor
DE protein-tyrosine kinase ErbB-1).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=84219729; PubMed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RT "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells."
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=95382957; PubMed=7654368;
RA Ilekis J.V., Stark B.C., Scoccia B.;
RT "Possible role of variant RNA transcripts in the regulation of
RT epidermal growth factor receptor expression in human placenta."
RL Mol. Reprod. Dev. 41:149-156(1995).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97078686; PubMed=8918811;
RA Reiter J.L., Maible N.J.;
RT "A 1.8 kb alternative transcript from the human epidermal growth
RT factor receptor gene encodes a truncated form of the receptor."
RL Nucleic Acids Res. 24:4050-4056(1996).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=97256547; PubMed=9103388;
RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;
RT "Expression of a truncated epidermal growth factor receptor-like
RT protein (TEGR) in ovarian cancer."
RL Gynecol. Oncol. 65:36-41(1997).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Placenta;
RX MEDLINE=21100872; PubMed=11161793;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGFR transcripts encoding truncated receptor
RT isoforms."
RL Genomics 71:1-20(2001).
```

```
RN [6]
RP SEQUENCE OF 575-687 FROM N.A.
RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Human and mouse alternative EGFR transcripts encoding only the
RT extracellular domain of the receptor."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; PubMed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: Gene
RT amplification and three related messenger RNA products in A431
RT cells."
RL Science 224:843-848(1984).
RN [8]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; PubMed=6330563;
RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells."
RL Nature 309:806-810(1984).
RN [9]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; PubMed=6093780;
RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells."
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [10]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=88217333; PubMed=3329716;
RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.D.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription."
RL Oncogene Res. 1:375-396(1987).
RN [11]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; PubMed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis."
RL J. Biol. Chem. 266:1746-1753(1991).
RN [12]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; PubMed=2991899;
RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene."
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [13]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [14]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; PubMed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA."
RL Nature 309:270-273(1984).
RN [15]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; PubMed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
```





```

FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE)
FT CARBOHYD 128 128 (BY SIMILARITY).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF2D2F5 CRC64;

Query Match 45.4%; Score 3083; DB 1; Length 1210;
Best Local Similarity 49.5%; Pred. No. 3.3e-155;
Matches 630; Conservative 164; Mismatches 362; Indels 116; Gaps 25;

QY 11 LLLALLPGAA--STQVCTGTDMKRLPASPEHLLMLRLHYQGVQVQGNLELTVLPTN 68
DB 11 LLLALLPGAA--STQVCTGTDMKRLPASPEHLLMLRLHYQGVQVQGNLELTVLPTN 68
QY 14 LTLTLCAGGALEEKVCGTSENRLTQGTDFEDHFLSLQRMVNCVVLGNLEITYVQRN 73
DB 14 LTLTLCAGGALEEKVCGTSENRLTQGTDFEDHFLSLQRMVNCVVLGNLEITYVQRN 73
QY 69 ASLSFLQDIQEVQVYLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128
DB 69 ASLSFLQDIQEVQVYLIHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDNGDPLNNTTP 128
QY 74 YDLSEFKTIOEVAGYVLIATNVTREPLENLQIIRGNALYENTYALAILSN----- 124
DB 74 YDLSEFKTIOEVAGYVLIATNVTREPLENLQIIRGNALYENTYALAILSN----- 124
QY 129 VTGASPGGLRELQSLRSLTEILKGGVLIQRNPOLCVQDTILWKDI-----FHKNQALATLI 184
DB 129 VTGASPGGLRELQSLRSLTEILKGGVLIQRNPOLCVQDTILWKDI-----FHKNQALATLI 184
QY 125 -YGTNRTGLRELPMNLQILIGAVRFSNNPLTCLNMDDTIQWRDVIQNVFMNSMDL---- 180
DB 125 -YGTNRTGLRELPMNLQILIGAVRFSNNPLTCLNMDDTIQWRDVIQNVFMNSMDL---- 180
QY 185 DTRNRACHPSCPMKSGSCWSESSEDCOSLRTVTCAGGCA-RCKGPLPTDCCHEQCAAG 243
DB 185 DTRNRACHPSCPMKSGSCWSESSEDCOSLRTVTCAGGCA-RCKGPLPTDCCHEQCAAG 243
QY 181 -QSHSSPCPKDPSFENGSCWGGGENCECKLTKIQAQCSHRCRGRSPSDCCNQCAAG 239
DB 181 -QSHSSPCPKDPSFENGSCWGGGENCECKLTKIQAQCSHRCRGRSPSDCCNQCAAG 239
QY 244 CTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303
DB 244 CTGPKHSDCLACLFHNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPNY 303
QY 240 CTGPRESCLVCQKQFODEATCKDTCPLMLYNTPTQMDVNFEGKYSFGATCVKCKPRN 299
DB 240 CTGPRESCLVCQKQFODEATCKDTCPLMLYNTPTQMDVNFEGKYSFGATCVKCKPRN 299
QY 304 LSTDVGSCTLVCPLNHNOVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFIGITELE 363
DB 304 LSTDVGSCTLVCPLNHNOVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFIGITELE 363
QY 300 VYTDHGSCVRACGPDYEV-EEDGIRKCKKCDGCRKVCNGIGIGEFK-DTLSINATNIK 357
DB 300 VYTDHGSCVRACGPDYEV-EEDGIRKCKKCDGCRKVCNGIGIGEFK-DTLSINATNIK 357
QY 364 -PAGCKIFGSLAFLPEFGDGPASNTAPLOEQVQVFTLEITGVLVISAWPDSLPL 422
DB 364 -PAGCKIFGSLAFLPEFGDGPASNTAPLOEQVQVFTLEITGVLVISAWPDSLPL 422
QY 358 HPKYCTALSGDLHIPLVAFKGSFTTLPDLPRELEILTKVKEITGFLLIQAMPNDTDL 417
DB 358 HPKYCTALSGDLHIPLVAFKGSFTTLPDLPRELEILTKVKEITGFLLIQAMPNDTDL 417
QY 423 SVFQNLQVIRGRILHNGAYSLTQGLISWGLRSLRELIGSLALIHNNHLCFCVHTVPW 482
DB 423 SVFQNLQVIRGRILHNGAYSLTQGLISWGLRSLRELIGSLALIHNNHLCFCVHTVPW 482
QY 418 HAFENLEIIRGRTKQHQGFSLAVVGLNITSLGRSLKEISDGDVVISGNRLCYANTINW 477
DB 418 HAFENLEIIRGRTKQHQGFSLAVVGLNITSLGRSLKEISDGDVVISGNRLCYANTINW 477
QY 483 DOLFENPHQALLHTANRDEDECVGEGSLACHOLCARGHCWGPGFTQCVNCSQFLRQECVE 542
DB 483 DOLFENPHQALLHTANRDEDECVGEGSLACHOLCARGHCWGPGFTQCVNCSQFLRQECVE 542
QY 478 KKLFTFTPNQTKIMNRAEKDKAVNVHVCNPLCSSEGCWGPDPFCVSCQNVSRGECVE 537
DB 478 KKLFTFTPNQTKIMNRAEKDKAVNVHVCNPLCSSEGCWGPDPFCVSCQNVSRGECVE 537
QY 543 ECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVACHYKDRPPFCVACRPS 602
DB 543 ECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCAVACHYKDRPPFCVACRPS 602
QY 538 KCNILEGEPRFVENSEICQHPCLPQAMNITCTGRGPDNCICQACHYIDGPHCVKTCPA 597
DB 538 KCNILEGEPRFVENSEICQHPCLPQAMNITCTGRGPDNCICQACHYIDGPHCVKTCPA 597
QY 603 GVKPDLSYMPIWKFDPESGACQPCPINCTHSCVDLDDKGPABQASPLTSTIQY-IXANS 661
DB 603 GVKPDLSYMPIWKFDPESGACQPCPINCTHSCVDLDDKGPABQASPLTSTIQY-IXANS 661
QY 598 GIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPLQGCCEWVPSGPKIPSIATGIVGL 656
DB 598 GIMGENNTL-VMKYADANNVCHLCHANCTYGCAGPLQGCCEWVPSGPKIPSIATGIVGL 656
QY 662 KFI-----GITELKRQKIRKYTWRRLLQETLEVEPLTPSGAMPNQAMRILKETELRK 716
DB 662 KFI-----GITELKRQKIRKYTWRRLLQETLEVEPLTPSGAMPNQAMRILKETELRK 716
```

```

DB 657 LFIVVVALGIGLFMRRRHIVRKRTRRLRLQERLEPLTPSGEAPNQAHLRLKETFEFK 716
QY 717 VKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSFKANKEILDEAVVMAGVSPV 776
DB 717 VKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSFKANKEILDEAVVMAGVSPV 776
DB 717 IKVLGSGAFGTVYKGLWIPDGENVKIPVAIKVLRENTSFKANKEILDEAVVMASVDNPHV 776
QY 777 SRLLGICLTSTVQLTQMPYGLCLLDHVRNRRGRGLSDQLLNMCMQIAKMSYLEVDRLV 836
DB 777 SRLLGICLTSTVQLTQMPYGLCLLDHVRNRRGRGLSDQLLNMCMQIAKMSYLEVDRLV 836
QY 837 HRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADCGKVPKIMMALESILRRPFTHQ 896
DB 837 HRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADCGKVPKIMMALESILRRPFTHQ 896
QY 837 HRDLAARNVVKTPQHVKITDFGLAKLLGAEKEKEYHABGGKVPKIMMALESILHRYTHQ 896
DB 837 HRDLAARNVVKTPQHVKITDFGLAKLLGAEKEKEYHABGGKVPKIMMALESILHRYTHQ 896
QY 897 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWMID 956
DB 897 SDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKCWMID 956
QY 897 SDVMSYGVTVWELMTFGSKPYDGIIPASDSSILEKGERLPQPPICTIDVYIMVVKCWMID 956
DB 897 SDVMSYGVTVWELMTFGSKPYDGIIPASDSSILEKGERLPQPPICTIDVYIMVVKCWMID 956
QY 957 SECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDDMDGLVDAE 1015
DB 957 SECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDDDDMDGLVDAE 1015
QY 957 ADSRPFRELILEFSKWARDPQRYLVIOGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD 1016
DB 957 ADSRPFRELILEFSKWARDPQRYLVIOGDERMHLPSPTDSNFYRALMDEEDMEDVVDAD 1016
QY 1016 EYLVPOQGFCDPDPAPGAGCMVHHRRSSSTRSGGDLTLGLEPSEEBAPRSPPLASEGA 1075
DB 1016 EYLVPOQGFCDPDPAPGAGCMVHHRRSSSTRSGGDLTLGLEPSEEBAPRSPPLASEGA 1075
QY 1017 EYLIPQGGFF-----NSPST-----SRTELLSLSA 1042
DB 1017 EYLIPQGGFF-----NSPST-----SRTELLSLSA 1042
QY 1076 GSDVFDGDLGMAAGLQSLTPDPLQRYSEDTPVLPSET--DGYVAPLTCSPQPEY 1133
DB 1076 GSDVFDGDLGMAAGLQSLTPDPLQRYSEDTPVLPSET--DGYVAPLTCSPQPEY 1133
QY 1043 TSN-----NSTVACIWRNGSCRKVEDAFQRYSDPTGAVTEIDIDDAFL-----PVPEY 1092
DB 1043 TSN-----NSTVACIWRNGSCRKVEDAFQRYSDPTGAVTEIDIDDAFL-----PVPEY 1092
QY 1134 VNQPDVRPQPPSPREGPLPAAPAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYL-T 1192
DB 1134 VNQPDVRPQPPSPREGPLPAAPAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYL-T 1192
QY 1093 VNQ-SVPKRPAGSVQNVYHNPQLHP-----APGRDLHYQN--PHSNAVGNPEYLT 1141
DB 1093 VNQ-SVPKRPAGSVQNVYHNPQLHP-----APGRDLHYQN--PHSNAVGNPEYLT 1141
QY 1193 PGGGAAPOPHPPPAPSPADNLYWDQ-----DP-----PERGAPSTFTKGTPT 1236
DB 1193 PGGGAAPOPHPPPAPSPADNLYWDQ-----DP-----PERGAPSTFTKGTPT 1236
QY 1142 AQ-----PTCLSSGFNSPALMTQKSHQMSLDNPDYQODFFPKPKNGIFKG-PT 1191
DB 1142 AQ-----PTCLSSGFNSPALMTQKSHQMSLDNPDYQODFFPKPKNGIFKG-PT 1191
QY 1237 AENPEYGLDVP 1248
DB 1192 AENAEYLRVAPP 1203

RESULT 6
ERR4_HUMAN
ID ERR4_HUMAN STANDARD; PRT; 1308 AA.
AC Q15303;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)
DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).
GN ERBB4 OR HER4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM JM-A).
RC TISSUE=Breast carcinoma;
RX MEDLINE=93189574; PubMed=8383326;
RA Plowman G.D., Culouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,
RA Foy L., Neubauer M.G., Shoyab M.;
RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the
RT epidermal growth factor receptor family."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).
RC TISSUE=Fetal brain;
RX MEDLINE=97476287; PubMed=9334263;
RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,
RA Klagsbrun M.;
```

RT "A novel juxtamembrane domain isoform of HER4/Erbb4. Isoform-specific  
 RT tissue distribution and differential processing in response to  
 RT phorbol ester".  
 RL J. Biol. Chem. 272:26761-26768(1997).  
 CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NTA. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND  
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -!- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY. IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; L07868; AAB59446.1; -.  
 CC HSSP; P11362; 1FGK.  
 CC Genew; HGNC:3432; ERBB4.  
 CC MIM; 600543; -.  
 CC InterPro; IPR000494; EGFR\_L\_domain.  
 CC InterPro; IPR000719; Euk\_Pkinase.  
 CC InterPro; IPR002174; Furin-like.  
 CC InterPro; IPR001245; Tyr\_pkinase.  
 CC InterPro; IPR004019; YLP\_motif.  
 CC Pfam; PF00069; pkinase; 1.  
 CC Pfam; PF00757; Furin-like; 1.  
 CC Pfam; PF01030; Recep\_L\_domain; 2.  
 CC Pfam; PF02757; YLP; 2.  
 CC ProDom; PD000001; Euk\_pkinase; 1.  
 CC SMART; SM00261; FU; 4.  
 CC SMART; SM00219; TyrKc; 1.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE; PSS0011; PROTEIN\_KINASE\_DOM; 1.  
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Alternative splicing.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 652 675 POTENTIAL.  
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 186 334 CYS-RICH.  
 FT DOMAIN 496 633 CYS-RICH.  
 FT DOMAIN 718 985 PROTEIN KINASE.  
 FT NP\_BIND 724 732 ATP (BY SIMILARITY).  
 FT BINDING 751 751 ATP (BY SIMILARITY).  
 FT ACT\_SITE 843 197 BY SIMILARITY.  
 FT DISULFID 189 197 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT DISULFID 213 221 BY SIMILARITY.  
 FT DISULFID 217 229 BY SIMILARITY.  
 FT DISULFID 230 238 BY SIMILARITY.

FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	181	181	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARSPLIC	626	648	NGPTSHDCIYVPMKTHSTLPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B)
FT	SEQUENCE	1308 AA;	146807 MM;	5E4AE0985D88761 CRC64;

Query Match 43.0%; Score 2925.5; DB 1; Length 1308;  
 Best Local Similarity 45.2%; Pred. No. 7.1e-147;  
 Matches 608; Conservative 175; Mismatches 385; Indels 177; Gaps 28;

QY	9	WGILLALLPPGAA----	STQVCTGTDMLRLPASPETHLDMLRLHLYGCGVOVGNLELT	64
DB	8	WYVSLVAAGTVQPSDSQSCAGTENKLSLSLEQQYRALRKYYENCEVVMGNLEITS	67	
QY	65	LPTNASLSFLQDIQEVQGVVLIAHNOVROVPLQRLRI	VRGTOLFEDNYALAVLNDGDP	124
DB	68	IEHRDLSFLRSVREVTGVVLVALNOFRYLPLENRI	IRGTCLYEDRYALAIFLNVRKDG	127
QY	125	NTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYQDTILKWDIFHKNQLALT	LI	184
DB	128	NF-----GLQELGLKLNTEILNGVYVDQNKFLCYADTIHMQDIVRNWPSNLT	LV	178
QY	185	DTNRSPACHPCSPMKGSCWGESSEDCQSLTRTVCAAGC-ARCKGPLPTDCHEQCAAG	243	
DB	179	STNGSSGCRCHKSCGTG-RCWGPTEHNHCQTLTRTVCAEQCDRCYGPVYSDCCHRECAGG	237	
QY	244	CTGPKHSDCLACLFHNHSGICELHCPALVYNTDTEFESMPNPEGRVTFGASCYVACP	PNY	303
DB	238	CSGPKDTCFACMNFDSGACVTCQPTVFYPTTFQLEHNFNAKVFACVKKCPHNF		297
QY	304	LSTDVGSCTLVCPLNHQEVTAEDGTQRCCKSPCARVCVGLGMQVIMKANSKEFIGITE	LE	363
DB	298	V-VDSSSCVRACPSSKMEV-EENGIMCKPCTDIPCACDGICTGSLMSAQTVDDSSNIDK		355
QY	364	FAGCKIFGSLAFLPESFGDDPASNTAPLOPOLQVFTLEBETGVLTYISAWPDSLPDLS	423	
DB	356	FINCTKINGNLIFLVTGIHGDYPNAIEADPEKLVNFRVREITGFLNIQSWPPNMTDFS	415	
QY	424	VFNQLQVIRGRIHNGAVSLTQGLGISWLRSGRELGLAIHHNTHLCFVHTVPWD	483	
DB	416	VFSNLVTIGRVLYSGLSLLILKQOQITLQFQSLKEISAGNIYITDNLCLYHHTINMT	475	
QY	484	QLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCWGPQTQCVCNCSQFLRGQCEV	EE	543

Db 476 TLSTINQIRIVIRDNRAENCAETAGMVCNHLCSGDCGMPGPDQCLSCRRFSRGRICIES 535  
Qy 544 CRVLOGLPREYVNAHCLPCHPEQOP-ONGSVTCFGEADOCVACAHYKDPFPFCVACBPS 602  
Db 536 CNLYDGEFREFENGSIQCEVDPQEKMDGLTCHGPGDCTKCSHFPGKDPNCVCEKCPD 595  
Qy 603 GVKPDLVMPWIKFDEBACQPCPINTCHSCVDLDDKGC-----PABQASPL 651  
Db 596 GLOGANSF--IPKYADPDRECHPCNPCTQCNGPTSHDCIYYPWTGHSTLPQHAR-TPL 652  
Qy 652 TSIQVIKANSXFI-----GIT--ELKRROQKIRKYTMRLLOETELVEPLTSPGAMPNOAQ 705  
Db 653 IAGVIL--GGILFIVLGLTFAVYVRRKS IKKRALRRFL-ETELVEPLTSGTAPNOAQ 709  
Qy 706 MRILKETELRKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 765  
Db 710 LRILKETELRKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 769  
Qy 766 YMGVGVSPYVRLGICLTSTVQLVLTOLMPYGCLLDHVRENRLGSLQDGLNMCWQIAK 825  
Db 770 LIMASMDHPHLVRLGLVCLSPITQILVQLMHPGCLLEYVHEKDNIGSQLLNMCVQIAK 829  
Qy 826 GMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMAL 885  
Db 830 GMYLEERLVRDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMAL 889  
Qy 886 ESILRRRTHQSDVWSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPPCTIDV 945  
Db 890 ECITHYKTHQSDVWSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPPCTIDV 949  
Qy 946 YIMVWKWIDSECRPFRELSEFSEMRDOPRVVIONED-LGPASPLDSTFYRSLLE 1004  
Db 950 YIMVWKWIDSECRPFRELSEFSEMRDOPRVVIONED-LGPASPLDSTFYRSLLE 1009  
Qy 1005 DDDMGDLVDAEYLVPOQGFCCPDAPAGAGGMVHRRHSRSGGGDTUTLGLPESEEA 1064  
Db 1010 EEDLEMDMAEYLVLP-QAFNIPPP-----IVTSRARDNSRS-----EIGHSPPPAYT 1057  
Qy 1065 PRS-----PLAP-SEGAGSDVFDGLGMAKGLQSL 1095  
Db 1058 PMSGNOFVYRDGGFAAEQGVSVYRAPRTSTIPEAPVAQGATAEIFDDSCNGTLRKPVP 1117  
Qy 1096 PTHDPSPLQYSDPTVPLPS-----ETDGVVAPLTCSPQPEYVNPQVVRPQPSPRE 1148  
Db 1118 HVQEDSSTORYSADPTVFAPERSGRGELDEGYWTPMRDKPQOYLNVP-----E 1168  
Qy 1149 GLPLAARPAAGATLERAKTLSPGKNGVVKOVFAFGAVENPEYLTPOGGAAPQHPHPPA-- 1206  
Db 1169 NPFVSR-----KNGDLQ-----ALDNPETHNASNG-----PPKAED 1200  
Qy 1207 -----FSPAFDNLVYNDQDPPERGA--PPSTFK 1232  
Db 1201 EYVNEPLYLNTFANTLGAEYLVKNILSMPEKAKKAFDNPDMNHSLLPPRSTLQHPDYLQ 1260  
Qy 1233 GTPT-----AENPEYL 1243  
Db 1261 EYSTKYFYKQNGRIRPIVAENPEYL 1285  
RESULT 7  
ERB4\_RAT  
ID\_ERB4\_RAT STANDARD; PRT; 1308 AA.  
AC Q62956; Q9Z2N7;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).  
GN ERB4 OR TYRO-2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=Heart;  
RX MEDLINE=98221155; PubMed=9553078;  
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,  
RA Marchionni M.A., Kelly R.A.;  
RT "Neuregulins promote survival and growth of cardiac myocytes.  
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult  
RT ventricular myocytes.";  
RL J. Biol. Chem. 273:10261-10269(1998).  
RN [2]  
RP SEQUENCE OF 848-901 FROM N.A.  
RC TISSUE=Sciatic nerve;  
RX MEDLINE=91225560; PubMed=2025425;  
RA Lai C., Lemke G.;  
RT "An extended family of protein-tyrosine kinase genes differentially  
RT expressed in the vertebrate nervous system.";  
RL Neuron 6:691-704(1991).  
RN [3]  
RP SEQUENCE OF 1031-1198 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuregulins and their putative receptors, ErbB2 and  
RT ErbB3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659(1997).  
CC -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS. NRG-  
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
CC NTK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
CC RECEPTORS (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING  
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS  
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE  
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND  
CC HEART.  
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC RESIDUES (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF041838; AAD08999.1; -;  
CC EMBL; U52531; AAC53051.1; -;  
CC HSP; P11362; IFGK  
CC InterPro; IPR000494; EGFR\_L\_domain.  
CC InterPro; IPR000719; Euk\_pkinase.  
CC InterPro; IPR002174; Furin-like.  
CC InterPro; IPR001245; Tyr\_pkinase.  
CC InterPro; IPR004019; YLP\_motif.  
CC Pfam; PF00757; Furin-like; 1.  
CC Pfam; PF00069; pkinase; 1.  
CC Pfam; PF01030; Recep\_L\_domain; 2.  
CC Pfam; PF02757; YLP; 2.  
CC PRINTS; PR00109; TYRKINASE.  
CC ProDom; PD000001; Euk\_pkinase; 1.  
CC SMART; SM00281; FU; 4.  
CC SMART; SM00219; TyrKc; 1.  
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
CC PROSITE; PS00111; PROTEIN KINASE DOM; 1.  
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 25

FT	CHAIN	26	1308	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT	DOMAIN	26	651	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	652	675	POTENTIAL.
FT	DOMAIN	676	1308	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	186	334	CYS-RICH.
FT	DOMAIN	496	633	CYS-RICH.
FT	DOMAIN	718	985	PROTEIN KINASE.
FT	NP BIND	724	732	ATP (BY SIMILARITY).
FT	BINDING	751	751	ATP (BY SIMILARITY).
FT	ACT SITE	843	843	BY SIMILARITY.
FT	DISULFID	189	197	BY SIMILARITY.
FT	DISULFID	193	205	BY SIMILARITY.
FT	DISULFID	213	221	BY SIMILARITY.
FT	DISULFID	217	229	BY SIMILARITY.
FT	DISULFID	230	238	BY SIMILARITY.
FT	DISULFID	234	246	BY SIMILARITY.
FT	DISULFID	249	258	BY SIMILARITY.
FT	DISULFID	262	289	BY SIMILARITY.
FT	DISULFID	293	304	BY SIMILARITY.
FT	DISULFID	308	323	BY SIMILARITY.
FT	DISULFID	326	330	BY SIMILARITY.
FT	DISULFID	503	512	BY SIMILARITY.
FT	DISULFID	507	520	BY SIMILARITY.
FT	DISULFID	523	532	BY SIMILARITY.
FT	DISULFID	536	552	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	559	577	BY SIMILARITY.
FT	DISULFID	580	589	BY SIMILARITY.
FT	DISULFID	593	614	BY SIMILARITY.
FT	DISULFID	617	625	BY SIMILARITY.
FT	DISULFID	621	633	BY SIMILARITY.
FT	MOD RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	138	138	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	174	174	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	253	253	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	358	358	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	410	410	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	473	473	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	548	548	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	620	620	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CONFLICT	1062	1062	S -> N (IN REF. 3).
FT	CONFLICT	1080	1082	PYT -> SYR (IN REF. 3).
SQ	SEQUENCE	1308 AA;	146957 MW;	D944BB0996A08B41 CRC64;

Query Match  
Best Local Similarity 44.9%; Pred. No. 6.7e-146;  
Matches 605; Conservative 182; Mismatches 395; Indels 164; Gaps 27;

Qy	1	MELA-ALCRWGLLL--ALLPGAASVCTGTDMLRLPASPETHDMLRLHLYGCGVQV	57
Db	1	MKLATGLWWSGLVAARTVQPSQSVCAGTENKLSLSLEQQYRALRKYENCEVVM	60
Qy	58	GNLELTLYPTNASLSFLQDIOEVOGYVLIHQNVRQVFLORLIRVRGTOLFEDNVALAVL	117
Db	61	GNLEITSIEHRDLISFLRESIREVTGYVLVALNQFYLEPLENLRIRGTLYEDRYALAI	120
Qy	118	DNGDPLNNTTPTVGASPGRLRELQRLSLTEILKGGVLIRQNPOLCYQDTILWKDIFHKNN	177
Db	121	LNRYKDGNF-----GLQELGLKNLTELINGGVYVDQNKFLCYADTIHWQDIVRNPW	171
Qy	178	QLATLTLDTNRSRACHPCSPMKSGRCSGESSEDCQSILTRTVACAGGC-ARCKGLPLTDCC	236
Db	172	PSNMTLVSTIGSSGCGRCHKSCG-RCWGPTENHCQTLTRTVCAEQDGRCYGYPVSDCC	230
Qy	237	HEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDFFSMNPENPEGRYTFGASCV	296
Db	231	HRECAAGSGPKDTCDFACMFNDNSGACVTCQCPOTFVYNPTTFOLEHFNFNKITYTGAFCV	290

Qy	297	TACPYNVLTSDVSGCTLVCPLHNOEVTAEADGTORCECKSPCARVCYVGLGMOYIKANSKF	356
Db	291	KKCHNFV-VDSSCVCRACSSKMEV--EENGIKMKCKPTDICTPRACDGIIGTSLMSAQTV	348
Qy	357	IGITELFAGCKKIFGSLAFILPESFDGDPASNTAPLOEQLOVETLEEITGYLYVISAMP	416
Db	349	DSSNIDKFINCTKINGNLIFLVTGIHGDVPYNAIDAIDPEKLVNFRVREITGFLNIQWP	408
Qy	417	DSLPLDSVFQNLQVIRGRIHLHGAYSILTLOGLGISWGLRSLRGLSGSLAIHNTHLCP	476
Db	409	PNMTDFSVFSLNLTIGRVLYSGLSLILKQGGITSLOFSLKSEISAGNIYITDNSLCY	468
Qy	477	VHTVPQLFRNPHQALHTANRDECEVGEGLACHQLCARGHCWGPQPTCVNCSQFLR	536
Db	469	YHTINWTTLSTVNRIVIRDNRAENCTAGMVCNHLCSNDGCMGPGPDQCLSCRRSR	528
Qy	537	GOECVBCRVLOGLPREYVNAHCLPCHPEQCP--ONGSVTCFGEADOCVCAHYKDPFF	595
Db	529	GKICIESCNLYDGEFRBFENGSI CVECDSQCEKMEGDLTCHGPGPDNCTKCSHFKQGN	588
Qy	596	CVARCPGVPDLSPYMPITWKFPPDEEGACQPCPINCTHSCVDLDDKGC-----PA	644
Db	589	CVEKCPDVLQANSF--IFKYADODRECHPCHPCTQCGNPTSHDCIYYPTWGHSTLUP	646
Qy	645	EQRASPLTSIQYIKANSKFIGI-----TELKRRQOKIRKYTMRLRLLQETELVEPLTPSG	698
Db	647	HAR--TPLIAGVI--GGFLIVIMALTFAVYVRKSIKKKRLARFL--ETELVEPLTPSG	702
Qy	699	AMPNQAQMRILKETELRKVKVGLSGAGFVYVGIWIPDGENVKIPIVAIKVLRNTPSKAN	758
Db	703	TAPNQAQLRILKETELRKVKVGLSGAGFVYVGIWVPEGETVKIPIVAIKILNETGPKAN	762
Qy	759	KEILDEAVVWAGVSPYVSRLLGLTSTVOLVTLMPYGCGLLDHVRNRCGLSODLLN	818
Db	763	VEFMEALIMASVDHPLVRLVLLGVCLSPITQLVTLMPHGCCLBYVHEHKNIGISQLLN	822
Qy	819	WCMTAKGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLRLDITDEYIADGGKV	878
Db	823	WCVOIAGMWELEERRLVHRDLAARNVLKSPNHVKITDFGLRLLEGEDEKEYNADGKM	882
Qy	879	PIKMALESILRRRPTHOSDWSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLQPP	938
Db	883	PIKMALECIHYRKFTQSDVWSYGVTVWELMTGKPKYDGIPTREIPDLLEKGERLQPP	942
Qy	939	PICTDIVYIMVKCMWIDSECRPRFRELVSFESFMRARDPQRFVUIONED-LGPASPLDST	997
Db	943	PICTDIVYIMVKCMWIDADSRPKFELAAEFMRARDPQRYLVIQDDRMKLPSPNDK	1002
Qy	998	FYRSLEDDDMGDLVDAAEYLVPOQGFCDPP-----APCAG	1034
Db	1003	FFQNLLDEEDLEDMWDAEYLVLP-QAENIPPIYTSRTRDSNRSEICHSPPAYTPMSG	1061
Qy	1035	GMVHRRSSSTRSGGDLTLGLEPSEEEAPRSPLAPSEGAGSDVFDGDLGMAAGLQOS	1094
Db	1062	SQFYVQGGFATQGG---MPMPYTATTSTIPEADVA--QGATAEMFDDSCCNGTLRKPVV	1116
Qy	1095	LPHTDPSPLORYSEDPVPLPS-----ETDGVVAPLTCSPQPEYVNPQDVRQPSPR	1147
Db	1117	PHVEDSSTQRYADPTVFAPERNPRAELDEEGYMTPMHDKPKQEYLNPE-----	1167
Qy	1148	EGPLPAARPAAGATLERAKTLSPGKNGVGVKDVFAFGGAVENPEYLTPOCGAAPQPHPPA-	1206
Db	1168	ENPVSRR-----KNGDLQ-----ALDNPYHSASSG-----PPKAE	1199
Qy	1207	-----PSPAFDNLYYWDQPPPERGA--PPSTF	1231
Db	1200	DEYVNEPLYNTFTNALGNAEYMKNSLLSVPEKAKKAFDNPYWNHSLPPRSTLQHPDYL	1259
Qy	1232	KGTP-----AENPEYL	1243
Db	1260	QESTKYFYKQNGRIRPIVAENPEYL	1285

```

RESULT# 8
ID_XMRK_XIPMA_STANDARD; PRT; 1167 AA.
AC_P13388;
DT_01-JAN-1990 (Rel. 13, Created)
DT_16-OCT-2001 (Rel. 40, Last sequence update)
DT_15-JUN-2002 (Rel. 41, Last annotation update)
DE_Melanoma receptor protein-tyrosine kinase precursor (BC 2.7.1.112).
GN_XMRK OR TU.
OS_Xiphophorus maculatus (Southern platyfish).
OC_Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC_Acinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC_Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;
OC_Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX_NCBI_TaxID=8083;
RN_[1]
RN_SEQUENCE FROM N.A.
RX_MEDLINE=90015140; PubMed=2797166;
RA_Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA_Telling A., Robertson S.M., Schartzl M.;
RT_"Novel putative receptor tyrosine kinase encoded by the melanoma-
RT_inducing Tu locus in Xiphophorus.";
RN_Nature 341:415-421(1989).
RN_[2]
RN_REVISION TO 515.
RA_Schartl M.;
RL_Submitted (JUL-2000) to the EMBL/GenBank/DBSJ databases.
CC_-!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC_-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC_tyrosine phosphate.
CC_-!- SUBCELLULAR LOCATION: Type I membrane protein.
CC_-!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC_-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
-----
CC_This SWISS-PROT entry is copyright. It is produced through a collaboration
CC_between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC_the European Bioinformatics Institute. There are no restrictions on its
CC_use by non-profit institutions as long as its content is in no way
CC_modified and this statement is not removed. Usage by and for commercial
CC_entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC_or send an email to license@isb-sib.ch).
-----
DR_EMBL; X16891; CAA34770.2; -.
DR_PIR; S06142; S06142.
DR_HSPG; P11362; IFKG.
DR_InterPro; IPR0000494; EGFR_L domain.
DR_InterPro; IPR000719; Euk_pkinase.
DR_InterPro; IPR002174; Furin-like.
DR_InterPro; IPR002290; Ser_thr_pkinase.
DR_InterPro; IPR001245; Tyr_pkinase.
DR_Pfam; PF00069; pkinase; 1.
DR_Pfam; PF00757; Furin-like; 1.
DR_Pfam; PF01030; Recep_L domain; 2.
DR_PRINTS; PR00109; TYRKINASE.
DR_ProDom; PD000001; Euk_pkinase; 1.
DR_SMART; SM00261; FU; 5.
DR_SMART; SM00220; S TKC; 1.
DR_SMART; SM00219; TyTKC; 1.
DR_PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR_PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR_PROSITE; PSS0011; PROTEIN KINASE DOM; 1.
KW_Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW_Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT_SIGNAL 1 25
FT_CHAIN 26 1167
FT_DOMAIN 26 642
FT_TRANSMEM 643 665
FT_DOMAIN 666 1167
FT_DOMAIN 710 977
FT_NE_BIND 716 724
FT_BINDING 743 743
FT_ACT_SITE 835 835
FT_DOMAIN 26 1167
FT_EXTRACELLULAR (POTENTIAL).
FT_POTENTIAL.
FT_CYTOPLASMIC (POTENTIAL).
FT_PROTEIN KINASE.
FT_ATP (BY SIMILARITY).
FT_ATP (BY SIMILARITY).
FT_BY SIMILARITY.

```



```
Db 530 GGRVASCNLLQGEPREAQVDRGCVCHQECVQLVQDLSLTCTCYGFGPANCSKSAHFQDGPQC 589
Qy 597 VARPSPGVKPDLSYMPIWKPFPDEGACQPCPINCINCHSCVDLDDKGGPAB-----QRASPL 651
Db 590 IPRCPHGILGDBGTL-IRKYADKMGQCPQCHQCTQCGSGPGLSGCRGDIVSHSLAVGL 648
Qy 652 TSIOYIKANSKFGITITELKRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRLIKE 711
Db 649 VSGLLITVIVALLIVLLRRRIK-RKRTIRCLLOEKELVEPLTPSGAQNQAFRLIKE 707
Qy 712 TELRKVKVGGAGFTYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGV 771
Db 708 TEFKDRVLGSGAGFTYKGLMNPQDGNIRIPVAIKVLEATSPKVNQEVLDAYVMASV 767
Qy 772 GSPVSELLGICITSTVOLTPMLPYCCLLDHVRNRRGLSGDOLLNMCQIAKMSYLE 831
Db 768 DHPHVCRLGICITSAVOLTPMLPYCCLLDYVRQHQERICGOWLLNMCVQIAKMNYLE 827
Qy 832 DVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKMALESILRR 891
Db 828 ERHLVHRDLAARNVLKSPNHVKITDFGLSKLLTADKEKYQADGGKVPKIKMALESILQW 887
Qy 892 RFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMVK 951
Db 888 TYHQSDVMSYGVTVWELMTFGSKPYDGIIPAKEIASVLENGERLPQPICTIEVYMIILK 947
Qy 952 CWMIDSECRPRFRELVESEFMRARDPQFVVIQNEGLGPASPLDSTFYRSLLEDMDGDL 1011
Db 948 CWMIDSECRPRFRELVESEFMRARDPQFVVIQNEGLGPASPLDSTFYRSLLEDMDGDL 1002
Qy 1012 VDAEYLVPOQGFPCDPAPGAGVGHRRSSSTRSGGGDLTLGLPESEEAAPRSLAP 1071
Db 1003 VDAEYLVPOQGFPCDPAPGAGVGHRRSSSTRSGGGDLTLGLPESEEAAPRSLAP 1025
Qy 1072 SEGAGSDVFGDLGMAAGLQSLPHTDPLQRYSDPPTV-PLPSTDGYVAPLTCSPQ 1130
Db 1026 PTGH-----PVRENSITLRNISDPQNALEKDLGDH----- 1056
Qy 1131 PEYVNOPDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLPNGKGVVKD 1177
Db 1057 -EYVNOPGSTSRSLDIYNPNVEDLTDGNGVSLSSQEAETNFSPEYLTNQNSL--- 1112
Qy 1178 VFAFGAVENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPPERPAPPTFKGTPTA 1237
Db 1113 PLYSSGSMDDPDY---QAG-----YQAAF-----LPQTGALTNGMFLPAA 1150
Qy 1238 ENPEYLG 1244
Db 1151 ENLEYLG 1157
```

```
RESULT 9
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
```

```
Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
[2]
SEQUENCE FROM N. A.
MEDLINE=90311312; PubMed=2164210;
RA Plozman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
[3]
SEQUENCE FROM N. A. (SHORT FORM).
RP TISSUE=Placenta;
RC MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M29366; AAA35790.1; -
CC EMBL; M34309; AAA35979.1; -
CC EMBL; S61953; AAB26935.1; -
CC PIR; A36223; A36223.
CC HSSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_Pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_Pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC ProDom; PD000001; Euk_Pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
CC PROSITE; PS00109; PROTEIN KINASE TYR; FALSE_NEG.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transfrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT POTENTIAL.
FT RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT PROTEIN KINASE.
FT ATP (BY SIMILARITY).
```

```

FT BINDING 742 742 ATP (BY SIMILARITY).
FT ACT_SITE 834 834 BY SIMILARITY.
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 141 183 SC -> GQPMVPSGLTPQPAQDWYLLDDPRLLTSLASSK
VPVTLAAV (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
E -> G (IN REF. 2).
E -> G (IN REF. 2).
SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;

Query Match 35.4%; Score 2407.5; DB 1; Length 1342;
Best Local Similarity 40.5%; Pred. No. 1.4e-119;
Matches 531; Conservative 189; Mismatches 455; Indels 137; Gaps 33;

Qy 10 GLLALLPPGAA--STQVCTGDMKRLPASPTHLDMLRHLYQGCQVQVQGNLEIYLTPT 67
Db 11 GLLPSLARGSEVGNQAQVCPGTLNGLSVTGAENQYQTLKLYERCEVVMGNLEIVLTGH 70

Qy 68 NASLSFLQDIOEVGYVLIANQVQVPLQRLRIVRGTLQDFEDYVALAVLDNGDPLNNTT 127
Db 71 NADLSFLQWIREVTGYVLIANNEFTLPLNLRVVRGTQVYDGKFAIPVM-----LNVNT 125

Qy 128 PVTGASPGLELQRLSLTEILKGGVLIQNPOLCYODTILWKDIFHKQNQLALTLDTN 187
Db 126 ----NSSHALQLRLTQLTEILSGVVIKKNLKLCHMDTIDWRDIVRDRD---AEIVVKD 178

Qy 188 RSRACHPCSPMKGRCWGESSEDCQLTRITVCAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPCHCEVKG-RCWGPGESEDCQLTKTICAPQCNGCHGFCFNPQCCHECAGCGSG 237

Qy 247 PKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYLSLT 306
Db 238 PQDTDCFACFHNDGAGVPRCPQPLVYNKLTFLQEPNPHTKYQYGGVGVCAVSCPHNEF-V 296

Qy 307 DVGSCTLVCLPHNOEVTAEQDQRCCKSPCARCYGLGQYIKANSKF--IGITBLE- 363
Db 297 DQTSQVACAPDKWEVD-KNGLKWCFCGGLCPKACEGTG-----SGSRFQTVDSWIDG 350

Qy 364 FAGCKKIFGSLAFPSDFGASNTAPLQEQVFTLEETIETGLYIYI SAWPDSLPLDLS 423
Db 351 FVNCTKILGNLDELITGLNGDPWHKIPALDPEKLNVTREITGYLNQSWPPHMFNS 410

```

```

Qy 424 VFQNLQVIRGRILHNGAYS-LTQGLGISWGLRLSLRELGSGLALIHNNHLCFVHTVPM 482
Db 411 VFSNLTTCGRSLYNRGPFLIMKNLNTSLGFRSLKETSAGRIYISANRQLCYHSLNW 470
Qy 483 DOLFNPQHALLHTA-NRPEDECVCEGLACHOLCARGHCWGPGTQVCNCSOFRLQOEVCV 541
Db 471 TKVLRGPTTEERLDIKNRPRDCAEAGKVCDDPLSCSGGCGWPGGQCLSCRNYSRGVCV 530
Qy 542 BECRVLQGLPREYVNNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARCP 601
Db 531 THCNPLNGEPREFAEAEFCFCHPECPQMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
Qy 602 SGVXPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAE-----QRASPLTS 653
Db 591 HGVLG--AKGPIYKYPDVQNECRPCHENCCTQCKGPELQDCLGQTLVLIGKTHLTWALT 648
Qy 654 IQYIKANSKFITELKRRQOKIR-KYTWRRLLQETELVEPLTPSGAMPNQAMRLKET 712
Db 649 IAGLVVIFPMILGGTFLYWRGRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIFKET 707
Qy 713 ELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKETLDEAYVMAGVG 772
Db 708 ELRKVLVLSGVGTGTVHKGWIPGESIKIPVICIIEDKSGRQSFQAVTDHMLAIGSD 767
Qy 773 SPYVRLIGICLTSTVQLVQLMPYGCILLDVHRENRLGSDLLNWCMIKAGMSYLED 832
Db 768 HAHVRLVLLGLCPGSSQLVTVQVPLGSLDHRQHRGALGPQLLLNWQVIAKGMYYLEE 827
Qy 833 VRLVHRDLAARNLVKSPNHVKITDPLGLARLLDIDETEHADGGKVPKIMMALESILRRR 892
Db 828 HGMVHRNLARNVLLKSPQVADFGVADLLPPDDKQLLYSEAKTPIKMMALESIFHGK 887
Qy 893 FTHOSDVMSYGVTVWELMTFCAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKC 952
Db 888 YTHOSDVMSYGVTVWELMTFCAKPYDGPAREIPDLLEKGERLPPOPICTIDVYIMVKC 947
Qy 953 WMIDECRPRELVSFSESRMARDPQFVVIQNEDLGPA---SPLDSTFYRSLLEDDDMG 1009
Db 948 WMIDENIRTPKELANEFTRMARDPPRYLVIKRES-GPGIAPGPEPHGLTNKKLEVELE 1006
Qy 1010 DLVDABEVLVPOQFFCPDPAPAGAGWVHRRHSSTRSGGDLTLGLEP-SEERAPRP 1068
Db 1007 PELDLDLDELEAD-----NLATTLGSLSLPVTGLNRRGQSOL 1047
Qy 1069 LAPSEGAGSDVFDGLGMAAGLQSLPTH-D-PSPLQRYSEDPTVPLP-----SETDGY 1121
Db 1048 LSPSGY-MPMNQNLGESCOESAVSGSSERCPRVSLH-----PMRGCGLASSESEGH 1100
Qy 1122 VA-----PLTCSQPE---YVNPQVVRPQPPSPREGP----- 1150
Db 1101 VTGSEAELEQKVSRCRSRSRSPRGRDSAYHSQRHSLLTPTVPLSPGLEEDVNGV 1160
Qy 1151 LPAARPAAGATLERAKTISP-GKNGV-----KDVFAFGAVENPEYLTPOGGAAPQHP 1203
Db 1161 MPDTHLGTPTSSREGTLSSVGLSSVLGTEEDED-----BEYEVNRRRRHSP-PHP 1211
Qy 1204 PPAFSPADNLYYWD-----QDPERCAPSTFKGTPTAENPEYL 1243
Db 1212 PRPSLEELGYEYMDVGSGLSASLGSTQSCPLHPVPIIPTAGTTPTDEDEYEM 1263

RESULT 10
ERR3_RAT
ID -ERR3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_taxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96096535; PubMed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
RT "Cloning of the rat ErbB3 cDNA and characterization of the  
recombinant protein.";  
RL Gene 165:279-284(1995).  
RN [2]  
RP REVISIONS TO 85; 513 AND 565.  
RA Hellyer N.J., Koland J.G.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 922-1097 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuroligins and their putative receptors, ErbB2 and  
ErbB3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659(1997).  
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
(POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
PHOSPHATIDYLINOSITOL 3-KINASE.  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
DR EMBL; U29339; AAC28498.2; -;  
DR EMBL; U52530; AAC53050.1; -;  
DR HSP; P11362; IFCK.  
DR InterPro; IPR000494; EGF\_R\_L domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR PRINTS; PR00109; Recep\_L domain; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Tyrosine-protein kinase; Multigene family; Receptor; Signal;  
Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 19  
FT CHAIN 20 1339  
FT DOMAIN 20 643  
FT TRANSMEM 644 662  
FT DOMAIN 663 1339  
FT DOMAIN 183 259  
FT DOMAIN 707 964  
FT NP\_BIND 713 721  
FT BINDING 740 740  
FT ACT\_SITE 832 832  
FT DISULFID 186 194  
BY SIMILARITY.

FT DISULFID 190 202 BY SIMILARITY.  
FT DISULFID 210 218 BY SIMILARITY.  
FT DISULFID 214 226 BY SIMILARITY.  
FT DISULFID 227 235 BY SIMILARITY.  
FT DISULFID 231 243 BY SIMILARITY.  
FT DISULFID 246 255 BY SIMILARITY.  
FT DISULFID 259 286 BY SIMILARITY.  
FT DISULFID 290 301 BY SIMILARITY.  
FT DISULFID 305 320 BY SIMILARITY.  
FT DISULFID 323 327 BY SIMILARITY.  
FT DISULFID 500 509 BY SIMILARITY.  
FT DISULFID 504 517 BY SIMILARITY.  
FT DISULFID 520 529 BY SIMILARITY.  
FT DISULFID 533 549 BY SIMILARITY.  
FT DISULFID 556 573 BY SIMILARITY.  
FT DISULFID 576 585 BY SIMILARITY.  
FT DISULFID 589 610 BY SIMILARITY.  
FT DISULFID 613 621 BY SIMILARITY.  
FT DISULFID 617 629 BY SIMILARITY.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 1028 1028 L -> P (IN REF. 3).  
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDFE1E CRC64;  
Query Match 34.1%; Score 2320.5; DB 1; Length 1339;  
Best Local Similarity 40.5%; Pred. No. 5.4e-115;  
Matches 520; Conservative 170; Mismatches 430; Indels 165; Gaps 35;  
QY 3 LAALCRWGLLLALLPPGAA---STQVCTGTDMKRLRPASPTHLDMLRLHYGCGVVOGN 59  
DB 7 LQVLC---FLLSLARGSEMGNQAVCPGTGLNGLSVTGDADNOVQTLKLYEKEVVMGN 62  
QY 60 LELYLPFNASLSFLQDIQEVQGVLIHAHNOVRQVPLQRLIVRQTLFEDNYALAVLDN 119  
DB 63 LEIVLTGHNADLSFLQWIREVTGTVLVANNEFSLVPLNLRVVRGTQVYDGFALFVM-- 120  
QY 120 GDPNNNTPTVTGASPGGLRELQLSLTEILKGGVLIQENPOLCYQDTILMKDIFHKNQL 179  
DB 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLCMDTIDWRDIVRVR-- 170  
QY 180 ALTLIDNRSRACHPCSPMKGSRGSCWSESSDCQSLTRTVCAAGC-ARCKGPLPLDCCHE 238  
DB 171 GAETVVKNGANGANCPPEVCKG-KWGGPGDDCQILTXTICAPQCNCRGCPGNPQCCHD 229  
QY 239 QCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPGRYTFGASCVTA 298  
DB 230 ECAGGCGSPQDTDCFACRRFNDSGACVPRCPPELVYNKLTQLEPNPHTKYQGGVCVAS 289  
QY 299 CPYNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGL--GMQYIKANSKF 356  
DB 290 CPHNFV-VDQTFVTRACPPDDQNEVD-KHGLKMCPECGGLCPKACEGTGSGRGYQTVDSN 347  
QY 357 IGITELBFAGCKIFGSLAFLPESPDGDPASNTAPLOPEQLQVFTLEITGLYLSIAMP 416  
DB 348 ID----GFVNCKILGNLDFLITGLNVDPWHKIPALDEKLVNFTVREITGLYNISQWP 403  
QY 417 DSLPDLVSFQNLQVIRGRLHNGAYS-LTLQGLGISWIGLSRLSRLGSLALIHNTHLIC 475  
DB 404 PHMFNFVSFNLTITGRSLYNRGFSLLIMKNLVNTSLGFRSLKEISAGRVVISANQQLC 463  
QY 476 FVHTVPWDQLFRNPQALLHTA-NRPEDECVGEGLACHQLCARGHCWGPFGTQCVNCSQF 534  
DB 464 YHSLNWTLLRGPSEERLDIKYDRPLGECIAEGKVCDPCLCSGGCGWGPQGCLSCRNY 523  
QY 535 LRQGECEVEECRVLQGLPREYVNRHCLPCHPECPQNGSVTCFGEADQCVACAHYKDP 594

```
Db 524 SREGVCTHCNLFQEGREFVHEAQCFSCHPECLPMEGTSTCNGSGSDACARCAFRDGP 583
Qy 595 FCVAPCPGVRDLSPYMIWKPDEGACQPCINCTHSCVDLDDKGPAPORASPLTSI 654
Db 584 HCVNCPHGILG--AKGPYIKYPAQNECRPCHECTGCGNPELQDCLGQ--AEVLMASK 639
Qy 655 QV-----IKANSKFTGITELKRRQOKIR-KYTMRRLLQETLVEPLTPSGAMPNQAO 705
Db 640 PHLVATVTVGLAVILMIGSFLYWRGRRIQKRAMRRYVLEGESEIEPLDPS-EKANKVL 698
Qy 706 MRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEA 765
Db 699 ARIFKETELRKVLGSGVGVGVHKGIMIPGESIKIPVCIKVIEDKSGRQSFQAVTDM 758
Qy 766 YVAGVGSFYVSRLLIGLICLTSTVLQVLTQMPYGLLDHVRNRLGSDLLNWCWQIAK 825
Db 759 LAVGSLDHAHIVRLGLCPGSSQLVQVPLPLGSLLDHVKQHRETLGPELLNMGVQIAK 818
Qy 826 GMSYLEDVRLHDLAARNVLKSPNHVKITDFGLARLLDIDETVHADGGKVPKKNWAL 885
Db 819 GMYLDEHSVHRDLARNVLSQVQVADFGVADLLPPDKQLLHSEAKTPIKNWAL 878
Qy 886 ESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDV 945
Db 879 ESIHFQKVTQSDVMSYGVYVWELMTFGAEPYAGRLAEIPDLLEKGERLAQPQICTIDV 938
Qy 946 YMIWKCMWIDSECPRELVSFSEMRARDQRFVIVQNEIDLGSPASPLDSTFYRSLLED 1005
Db 939 YMVWKCMWIDENIRPTFKELANETFRWARDPRYLVIKRAS-GRGTP--PAAEPSVLTT 995
Qy 1006 DMGDLVDAEEVLVPOQGFCCPDAPGAGGMVHRRSSSTRSGGDLTLGLEPSEE--- 1062
Db 996 KEL-----QEALEFEL-----DLDLDEAEEGEA 1021
Qy 1063 -----EAPRSLPAPSEG-----AGSDVFDGDLGMAAKGLQSLPT 1097
Db 1022 TSLGSALSPLTCTLRPGSQSLSPSSGYMPPMNSGLGEACLSAVLGGREFSRPISL 1081
Qy 1098 HDPSPLOYSEDPVPLPSETDGVY-----APL-----TC-----SPOPE-----YVNPQ 1137
Db 1082 H-PIPRGR-----PASESEGHVTGSEAELEQKSVCSRSRSPRPRGDSAYHSQR 1133
Qy 1138 DVROPSPSPREGP-----LPAARPAGATLERAKTLP-KGNQVW-----KQVF 1179
Db 1134 HSLTPTVPLSPGLLEEDGNGYVMPDTHLRGASSREGTSLSVGLSSVLGTREEDED--- 1191
Qy 1180 AFGGAVENPEYLTPOGGAAPQPHPP 1204
Db 1192 -----EEVEYNNRRKRGSP-PRPP 1209
RESULT 11
ID EGFR_DROME STANDARD; PRT: 1426 AA.
AC P04412: O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).
RP MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
```

```
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN
RP REVISIONS.
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION
RP ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Blannkocch C., Balow D.L.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cui S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupek M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN
RP SEQUENCE OF 959-1078 FROM N.A.
```



```
Db 543 RPAIQKEPEQKWNENLRADLCENKGTICSDQNCDCGAGTDCQCTCKNFNPGTC 602
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPFCVARC 600
Db 603 IADCGVISNAYK--FDNRCKICHPECR-----TCNGAGADHCQECVHVRDQHCVCSEC 654
QY 601 P-----SGVK-----PDL 608
Db 655 PKNKYNDRGVCRCHATCGCTGPKDTIGIGACTTCNLAIINNDATVVRCKLLKDDKCPD- 713
QY 609 SYMPIWKFPDEEGACQP-----CPI-----NCTH-----632
Db 714 GY--FWYVHPHQESLKLPLAGRAVCRKCHPLCELCNTYGYHEQVCSKTHYRREQCBT 771
QY 633 -----SC-----VDLDDKG-----641
Db 772 ECPADHYDEQRECFORHPECNCGTGPAGDDCKSCRNPKFLFDANETGYPVNSTWPNCTS 831
QY 642 -CPAEQR-----ASPLTSIQY---IKANSKFI-----GITEL 669
Db 832 KCPLMRHVNYOYTAIGPYCAASPPRSSKITANLDVNMIFITGAVLVPTICILCVVTYI 891
QY 670 KRRQOKIRKYT--MRRLLQETELVEPLTPSGAMPNQAOMRILKETELRKVKVLGSGAGFT 727
Db 892 CRQOKAKKETVMTMALSGCEDSEPLRPSNIGANLCKLRIVKDAELRKGVLGNGAFGR 951
QY 728 VYKGIWIPDGENVKIPVAIKVURENTSPKANKEILDEAYVMAGVSPYVSRLLGLCLTST 787
Db 952 VYKGVVPEGENVKIPVAIKELLKSTGAESSEFLREAYIMASEHVNLKLLAVCMSSQ 1011
QY 788 VOLVTQMPYGCGLDVRHNRGLCSQDLNWMQIAKMSVLELDVRLVHRDLAARNVLV 847
Db 1012 MMLITQMLPLGCLLDYVRRNRDKIGSKALLNWMSTQIAKMSYLEKRLVHRDLAARNVLV 1071
QY 848 KSPNHVKITDFGLARLLDIDETEHADGGKVKPIKMALESILRRFTHQSDVMSYGVTVW 907
Db 1072 QTPSLVKITDFGLAKLLSSDSNEYKAAGKPKIKWALECIRNVRTSKSDVWAFGVTVW 1131
QY 908 ELMTFGAKYDGIIPAREIPDLEKGERLPQPPICTIDVYIMVWKCMIDSECRPRFREL 967
Db 1132 BLLTFQGRPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCWHLDAAMRPTFKLT 1191
QY 968 SEFSRMRDPRFVVIQNEGLG--PASPLDSTFYRSLLEDD---DMGLDVAEAYLVPOQ 1022
Db 1192 TVFAEFADPGRYLAIPGDKFTRLPA-----YTSQDEKOLIRKLAPTTDGESEAIKPD 1244
QY 1023 GFFCPDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEERAP-----RGLAPSEAG 1076
Db 1245 DYLQPKAAPGPS-----HRTDCT-----DEMPLKNRYCKDPSNKNSTG 1283
QY 1077 SDVFDG----DLGMGAAGLQSLPHDPSPLORYSEDPTVPPLPSETDGYVAPLTCSPQPEY 1133
Db 1284 DDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCQPGNN 1321
QY 1134 VNQPDVRPQPPSPREGPLPAARPGATLERAKTLSPGKNGVVKVAFAGGAVENPEYL-- 1191
Db 1322 NNNNN-----NPNQNNNAAGVAAGYM-----DLIGVPVSDNPEYLLN 1360
QY 1192 --TPQGGAAPOPH-----PPAFSP-AFDNLYYMD 1218
Db 1361 AQLGVGESPIPTQIGIPVNGGPGTMEVKVPMGSEPSTSSDHEYND 1408

RESULT 12
ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
```

```
OS Avian leukosis virus.
OC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.
RN NCBI_TaxID=11864;
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
CC PIR; A00643; TVCHLV.
CC PIR; B00643; TVFVLV.
CC HSP; P11362; 1FGK.
CC InterPro; IPR000719; Euk_kinase.
CC InterPro; IPR001245; Tyr_kinase.
CC Pfam; PF00069; kinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC ProDom; PD000001; Euk_kinase; 1.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
CC Transfaser; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC KW DOMAIN 132 399 PROTEIN KINASE.
CC FT NP BIND 138 146 ATP (BY SIMILARITY).
CC FT BINDING 165 165 ATP (BY SIMILARITY).
CC FT ACT SITE 257 257 BY SIMILARITY.
CC SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.1%; Score 1706.5; DB 1; Length 634;
Best Local Similarity 51.6%; Pred. No. 5.2e-83;
Matches 364; Conservative 73; Mismatches 142; Indels 127; Gaps 16;

QY 587 CAHYKDPFPFCVARGSPGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQ 646
Db 3 CAHFDGPHCVKACPAAGVLENDTL-VWKYADANAVCOLCHPNCCTRGCKPGLEGCP--- 58
QY 647 RASPLTSIQYTKANSKF-----IGITELKRRQOKIRKYTVRRLLQETELVEPLTPSGAM 700
Db 59 NGSKTPSIAGVGVGLLCLLVVVGIGLGLYLRHRRHVRKRTLRLRLQERELVEPLTPSGEA 118
QY 701 PNOAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVURENTSPKANKE 760
Db 119 PNOAHLRLKETEPKVKVLGSGAGFTVYKGLWIPGEKVKIPVAIKELREATSPKANKE 178
QY 761 ILDEAYVMAGVSPYVSRLLGLCLTSTVQLVTQMPYGCGLDVRHNRGLCSQDLNWM 820
Db 179 ILDEAYVMASVDNPHVCRLLGLCLTSTVQLITQMPYGCGLDVRHNRGLCSQDLNWM 238
QY 821 MQIAKMSYLELDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETEHADGGKVP 880
```

```

Db      239 VOIAGMNYLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLGADKEYHAEGKVPPI 298
Qy      881 KMALESILRRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPPI 940
Db      299 KMALESILHRIYTHQSDVWSYGVTVWELMTGSKPYDGPASEISSVLEKGERLPQPPPI 358
Qy      941 CTIDVYMLVWKCMWIDSECRPFRELVSFSMARDPORFVVIQ-NEDLGPASPLDSTFY 999
Db      359 CTIDVYMLVWKCMWIDADSRPFRELIAEFSKWARDPPRYLVIOGDERMHLPSPTDSKPY 418
Qy      1000 RSLLEDDDDMDGLVDAAEYLVLPQGFPCDPAPACAGMWHRRSSSTRSGGGDLTLGLEP 1059
Db      419 RTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy      1060 SEEEAPRSP-----APSEGASDVDFDGLGMAKGLQSLTHDPSPLORSEDTVPL 1114
Db      450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHFVREDSFVQRYSSDPTGNF 496
Qy      1115 PSET--DGTVAPLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPGATLERAKTILSPGKN 1172
Db      497 LEESIDDDGFL-----PAPEYVNO--LMPKKSTAM----- 524
Qy      1173 GVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPPAFAPFONLYYM 1217
Db      528 QIYNNISLTAISKLPMSDRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPYW 579
Qy      1218 DQ-----DPPE-----RGAPSTFKGTPTAENPEYLGIDVP 1248
Db      580 IQSGNHQINLNDPQYQDFLPNETKPNGLLVKVPAAENPEYLRVAAP 625

```

## RESULT 13

```

ERBB AVIER
ID ERBB AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ES4).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]_TaxID=79685;
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RT gene family."
RL Cell 35:71-78 (1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328658;
RA Debuire B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
RT new type of oncogene."
RL Science 224:1456-1459 (1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; K020006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1;
DR PIR; A00644; TVYUH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match 24.4%; Score 1660; DB 1; Length 604;
Best Local Similarity 51.4%; Pred. No. 1.4e-80;
Matches 354; Conservative 70; Mismatches 135; Indels 130; Gaps 15;

Qy 587 CAHYKDPFPCVARCPGSKVPDLSPYMPKPFDEEGACQPCPINCTSCVDLDDKGCPAEQ 646
Db 3 CAHFIDGHCVCACAGVLGENDTL-VKAYADANAVCQLCHPNCTCKGPGLEGCP--- 58
Qy 647 RASPLTSIQYIKANSKF-----IGITELKRRQOKIRKYTMRLLOQTELVEPLTPSGAM 700
Db 59 NGSKTPSIAAGVVGGLLVVGLGIGLYLRRRHIVRKETLRLLQERELVEPLTPSGEA 118
Qy 701 PNAQWRILKETELRKVKVLGSGAGTGYKGIWIDGENVKIPVAIKVRENTSPKANKE 760
Db 119 PNAQHLRLKETEFKKVKVLGSGAGFTYKGLWIPEGEKVKIPVAIKELREATSPKANKE 178
Qy 761 ILDEAYVMAGVGSPPVSRLLGLCLSTVOLTPMLPYGCLLDHVRNRCRLGSGDNLNWC 820
Db 179 ILDEAYVMASVDNPHVCRLLGLCLSTVOLITQMLPYGCLLDYIREHKONIGSQYLLNWC 238
Qy 821 MQIAGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKPI 880
Db 239 VQIAGMNYLEERRLVHRDLAARNVLKTPQHVKITDFGLAKLGADKEYHAEGKVPPI 298
Qy 881 KMALESILRRRFTHQSDVWSYGVTVWELMTGAKPYDGPAREIPDLLEKGERLPQPPPI 940
Db 299 KMALESILHRIYTHQSDVWSYGVTVWELMTGSKPYDGPASEISSVLEKGERLPQPPPI 358
Qy 941 CTIDVYMLVWKCMWIDSECRPFRELVSFSMARDPORFVVIQ-NEDLGPASPLDSTFY 999
Db 359 CTIDVYMLVWKCMWIDADSRPFRELIAEFSKWARDPPRYLVIOGDERMHLPSPTDSKPY 418
Qy 1000 RSLLEDDDDMDGLVDAAEYLVLPQGFPCDPAPACAGMWHRRSSSTRSGGGDLTLGLEP 1059
Db 419 RTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1060 SEEEAPRSP-----APSEGASDVDFDGLGMAKGLQSLTHDPSPLORSEDTVPL 1114
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHFVREDSFVQRYSSDPTGNF 496
Qy 1115 PSET--DGTVAPLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPGATLERAKTILSPGKN 1172
Db 497 LEESIDDDGFL-----PAPEYVNO--LMPKKSTAM----- 524
Qy 1173 GVVKDVF-----AFGAVENPEYLTPOGGAAPQPHPPPAFAPFONLYYM 1217

```



```

Db 525 -VQNOIYNFISLTALSKLPMSRYQNSHSTAVDNPEYL-----NTNQSPKATVPES 575
QY 1214 LYWDQDPPERGAPSTFKGTPTAENPEY 1242
Db 576 SPYMIQSGNHQ-----INLNDPDY 594

RESULT 14
ERBB_AVIEU
ID ERBB_AVIEU STANDARD; PRT; 540 AA.
AC P11273;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain ts167).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=103898;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=87064458; PubMed=2878364;
RX Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;
RT "A single amino acid substitution in v-erbB confers a thermolabile
RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid
RT cells.;"
RL Mol. Cell. Biol. 6:1751-1759(1986).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M13179; AAA42401.1; -
DR PIR; A25231; TVFVEB.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).
SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA06B65D CRC64;

Query Match 23.4%; Score 1587; DB 1; Length 540;
Best Local Similarity 54.2%; Pred. No. 8.5e-77;
Matches 335; Conservative 63; Mismatches 126; Indels 94; Gaps 13;

QY 587 CAHYKDPFCVARGSPGVKPDLSYMPIWKFPDEGACOPCPINCTHSCVDLDDKGCRAEQ 646
Db 3 CAHFIDGPHCVKACPAAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGGLGSCP--- 58
QY 647 RASPLTSIQYKANSKF-----IGITELKRQOKIRKYWRLLQETELVEPLTPSGAM 700
Db 59 NGSKTPSIAAGVGGGLLCLVVGVLGILGLYLRHVRKTRRLRLQLRELVEPLTPSGEA 118
QY 701 PNOQMRLKETELRKVKVLGSGAFGTVYKGIWIPDGENVKVIPVAIKVLRNTSPKANKE 760

```

```

Db 1119 PNOAHLRLKETEFKKVKVGLFGAGFTVYKGLWIPGEKVTIPVAIKELREATSPKANKE 178
QY 761 ILDEAYVMAGVSPYVRLGLICLTSTVOLVTOLMPYGCCLLDHVRNRCRLGSSODLLNWC 820
Db 179 ILDEAYVMASVDNPHVCRLLGLICLTSTVOLITQLMPYGCCLDYIRHKONIGSQYLLNWC 238
QY 821 MOIAKMSYLEDVRLVHRDLAARNVLKSPNNHKITDFGLARLLDITETSYHADGKVP 880
Db 239 VOIAKGMNLEERHMHVRDLAARNVLKTPQHKITDFGLAQKLGADKEKYEHAEGKVP 298
QY 881 KMALESILRRPFTHQSDWMSYGVTVWELMTGCAKPYDIPAREIPDLLEKGERLLPOPP 940
Db 299 KMALESILHRIYTHQSDWMSYGVTVWELMTGSKPYDIPASEISSVLEKGERLLPOPP 358
QY 941 CTIDVYMWKCMWIDSECRPRFELVSEFSMARDPQRFVVIQ-NEDLGPASPLDSTFY 999
Db 359 CTIDVYMWKCMWSDARSFKRELIAEFKWARDPPRYLVIOGDERVHLSPTDSKFY 418
QY 1000 RSLDEDDMDGLVDAAEYLVPOQGFPCDPAPGAGGMVHRRHSSTRSGGDLTLGLEP 1059
Db 419 RTLMEEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
QY 1060 SEERAPRSP-----APSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPTVPL 1114
Db 450 -----SRTFLSSLSATSNSATNCIDRNG-----H----- 476
QY 1115 PSETDGYVAPLTCSPQPEVNPQVPRPQPSREGPLPAARAGAT-LERAKTLSPGKNG 1173
Db 477 PVREDGFL-----PAPEYVQ--LMPKKPSTAMVQNOIYNVYISLTATSKLPIDSRYN- 527
QY 1174 VKQVFAFGGAVENPEYL 1191
Db 528 -----SHSTAVDNPEYL 539

RESULT 15
EGFR_CHICK
ID EGFR_CHICK STANDARD; PRT; 703 AA.
AC P13387;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER
DE (Fragment)).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha.;"
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds  
(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-369-383-12

Perfect score: 6815

Sequence: 1 MELAALCRWGLLLALLPRGA.....TFKGTPTAENPEYLGLDVPV 1255

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6152	90.3	1259	6	O18735 canis famil
2	3096	45.4	1209	11	Q9GX70 rattus norv
3	3067	45.0	1210	11	Q9EP98 mus musculu
4	2692	39.5	1165	13	Q3YH40 xiphophorus
5	2645.5	38.8	1137	13	Q9W6F6 gallus gall
6	2264	33.2	1328	13	P79754 fugu rubrip
7	2010.5	29.5	1433	5	Q9BIH9 anopheles g
8	1871	27.5	419	4	Q9UK79 homo sapien
9	1739	25.5	367	11	Q8R2X1 mus musculu
10	1720	25.2	729	15	Q86712 avian rous-
11	1718	25.2	567	15	Q86714 avian rous-
12	1697.5	24.9	412	4	Q8WV0 homo sapien
13	1662.5	24.4	962	15	Q64895 avian eryth
14	1645	24.1	545	15	Q85468 avian eryth
15	1458.5	21.4	655	11	Q9WVF5 mus musculu
16	1442.5	21.2	643	11	Q9ERV6 mus musculu

17	1250	18.3	1193	5	Q9Y1X8
18	1178.5	17.3	1368	5	Q23821
19	1161	17.0	1717	5	Q26566
20	1084	15.9	527	13	Q90836
21	953.5	14.0	478	11	Q9ESE0
22	906	13.3	165	4	Q14256
23	900.5	13.2	599	13	Q9FSH2
24	887	13.0	176	11	Q923V5
25	806.5	11.8	346	13	P11776
26	778	11.4	435	5	Q8SZW1
27	754.5	11.1	311	13	Q9Y162
28	734.5	10.8	1362	13	Q9PVZ4
29	734	10.8	331	4	Q9BUD7
30	732	10.7	1671	5	Q9NJV5
31	723	10.6	149	6	Q9BG66
32	690	10.1	1418	13	Q93457
33	675.5	9.9	1368	13	Q8UW85
34	659.5	9.7	1369	13	Q8UW86
35	659.5	9.7	1472	5	Q9U5A8
36	659	9.7	1358	13	Q73798
37	657	9.6	1412	13	Q8UW84
38	644.5	9.5	1418	13	Q8UW83
39	625	9.2	1371	11	Q9QVW4
40	624	9.2	1245	13	Q9YGH8
41	614.5	9.0	2144	5	Q9VD94
42	589	8.6	935	4	Q96L35
43	587.5	8.6	1036	4	Q07912
44	587.5	8.6	1055	11	O54967
45	586.5	8.6	1091	4	Q9UMQ4

#### ALIGNMENTS

#### RESULT 1

O18735 PRELIMINARY; PRT; 1259 AA.  
AC O18735  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Erbb-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]\_TaxID=9615;  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erbb-2 from canine mammary gland.";  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1;  
DR HSSP; P11362; 1FGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Q9Y1X8 ephydatia f  
Q23821 caenorhabdi  
Q26566 schistosoma  
Q90836 gallus gall  
Q9ESE0 rattus norv  
Q14256 homo sapien  
Q9FSH2 gallus gall  
Q923V5 rattus norv  
P11776 xiphophorus  
Q8SZW1 drosophila  
Q9Y162 xiphophorus  
Q9PVZ4 xenopus lae  
Q9BUD7 homo sapien  
Q9NJV5 biophalari  
Q9BG66 oryctolagus  
Q93457 scophthalmu  
Q8UW85 paralichthy  
Q8UW86 paralichthy  
Q9U5A8 bombyx mori  
Q73798 xenopus lae  
Q8UW84 paralichthy  
Q8UW83 paralichthy  
Q9QVW4 rattus sp.  
Q9YGH8 scophthalmu  
Q9VD94 drosophila  
Q96L35 homo sapien  
Q07912 homo sapien  
O54967 mus musculu  
Q9UMQ4 homo sapien

Query Match	90.3%;	Score 6152;	DB 6;	Length 1259;
Best Local Similarity	90.0%;	Pred. No. 0;		
Matches 1134;	Conservative 45;	Mismatches 75;	Indels 6;	Gaps 2;

Qy	1	MELAAALCRWGLLLALLPPGAAS	TOVCTG	TDMLRLPAS	PTHLDMLRHL	YOGCOV	QGNL	60
Db	1	MELAAWCRWGLLLALLPFGAAG	TQVCTG	TDMLRLPAS	PTHLDMLRHL	YOGCOV	QGNL	60
Qy	61	ELTYLPTNASLFLQDIOEVGY	VLIAH	NQVRQVPL	QRLRIVR	GTQTF	FEDNYALAV	LDNG
Db	61	ELTYLPANASLFLQDIOEVGY	VLIAH	SQVRQIP	QLQRLRIV	RTQTF	FEDNYALAV	LDNG
Qy	121	DPLNNTTPTVGASPGGLRE	LQLSR	TEILKGG	VLIQRN	PLCYQ	DTILWKD	IFHKNQOLA
Db	121	DPEGGIPAPGAAGGLRE	LQLSR	TEILKGG	VLIQRS	PLQCH	ODTILWKD	VFHKNQOLA
Qy	181	LTLIDTNRSRACHPCSP	MCKSGR	CMWGSSE	DQCSL	TRTV	CAGGCAR	CKGPLTDCCH
Db	181	LTLIDTNRFSA	CPSPACKDA	HCWASS	GDQCSL	TRTV	CAGGCAR	CKGPOPTDCCH
Qy	241	AAGCTGPKHSDCLAC	LHFNHSGI	CELHCP	ALVTY	NTDT	FESMPN	PEGRTYFGASCV
Db	241	AAGCTGPKHSDCLAC	LHFNHSGI	CELHCP	ALVTY	NTDT	FESMPN	PEGRTYFGASCV
Qy	301	YNYLSTDVGSCTLC	VLPHNQV	BTADGT	QRCCKS	KPCAR	VCYGLG	WQYIKANSK
Db	301	YNYLSTDVGSCTLC	VLPLNNQV	BTADGT	QRCCKS	KPCAR	VCYGLG	MEHLREVRAV
Qy	361	ELEFAGCKT	FGSLAF	IPES	FDGDP	ASNTA	POYIKANS	KFTIGITEL
Db	361	IQEFAGCKT	FGSLAF	IPES	FDGDP	ASNTA	PLQPEQ	LRVFEALBEITGY
Qy	421	DLVSFQNLVIRGR	ILHNGAY	SLTLQ	LGLIS	WLG	LSRL	ELGSLALIHNT
Db	421	NLSVFQNLVIRGR	VLHDGAY	SLTLQ	LGLIS	WLG	LSRL	ELGSLALIHNR
Qy	481	PWDQLFRNP	QHALHTANR	PEDE	CVGEG	LACHQ	I	CARGHCWG
Db	481	PWDQLFRNP	QHALHSANR	PEEE	CVGEG	LACYP	-	CAHGH
Qy	541	VEECRV	LQGLPRE	VYNAR	HC	LPCH	PE	CPQNGSVTCF
Db	540	VEECRV	LQGLPRE	VYKDY	CLP	CH	SE	CPQNGSVTCF
Qy	601	PSGVK	PDLSYMPI	WKFP	DEE	GACQ	PCP	INCTHSCVDL
Db	600	PSGVK	PDLSFMP	INWF	ADE	EGT	QC	PCP
Qy	661	ILLVV	LGVVFGI	ILKRR	QOKIR	KYTR	MRLLQ	ETELVE
Db	660	ILLAV	VGLVLGI	ILKRR	QOKIR	KYTR	MRLLQ	ETELVE
Qy	721	RKVKVL	SGSAG	FTVY	KG	WIP	DGEN	VKIPVAI
Db	720	RKVKVL	SGSAG	FTVY	KG	WIP	DGEN	VKIPVAI
Qy	781	YVSR	LLGICLT	STVOL	TP	Q	LM	PYGCLLDH
Db	780	YVSR	LLGICLT	STVOL	TP	Q	LM	PYGCLLDH
Qy	841	LVR	DLAARN	V	LKSP	NHVKIT	DFGL	ARLLDIDET
Db	840	LVR	DLAARN	V	LKSP	NHVKIT	DFGL	ARLLDIDET
Qy	901	HQSD	W	SV	GV	TW	ELMT	FGAKPYDGI
Db	900	HQSD	W	SV	GV	TW	ELMT	FGAKPYDGI
Qy	961	IDSE	CR	PR	FR	EL	VSE	FSR
Db	960	IDSE	CR	PR	FR	EL	VSE	FSR
Qy	1021	ESYL	V	PO	Q	GF	CP	DPAPAG

[illegible]



```
KW ATP-binding; Receptor; Transferase.
SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 45.0%; Score 3067; DB 11; Length 1210;
Best Local Similarity 49.4%; Pred. No. 1.9e-222;
Matches 629; Conservative 164; Mismatches 368; Indels 112; Gaps 25;

QY 11 LLLALLPPGAA--STQVCTGTDMKRLRLPASPTHLDMLRLHLYQGQVVGQNLLEYLPTN 68
DB 14 LLLALLPPGAA--STQVCTGTDMKRLRLPASPTHLDMLRLHLYQGQVVGQNLLEYLPTN 68
QY 69 ASLSFLQDIQEOGVVLIHNVQVPLQRLAIRVGTQLFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLKTIQEVAGVLLIATNVERIPLENLQIIRGNALYENTYALATLSN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYODTILWKDI---FHKNOLALTLI 184
DB 125 -YGNRTGLRELPMKLNQELIIGAVRFSNNPILCMNDIIQWRDIVQNVFMSMDL--- 180
QY 185 DTRNSRACHPCSPMCKGSRMGESSEDCQSLTRTVACGGCA-RCKGPLPTDCCHQCAAG 243
DB 181 -QSHPSCKCDPSCPNCGSCWGGGECQKLTIIICAQCSHRCHGRSPSCCHNQCAAG 239
QY 244 CTGPKHSCLACLPHNHSIGIELCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNY 303
DB 240 CTGPRESCLVCQRFQDEATCKDTCPPLMLYNTTYQMDVNPPEKYSFGATCVKCKPRY 299
QY 304 LSTDVGSCTLCPLHNQVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGITELE 363
DB 300 VTDHGSCVRAGPDYEV-FEDGIRKCKKCDGPKRKYCNIGIGIEFK-DTLSINATNIK 357
QY 364 -FAGCKKIFGSLAFIPESFDGDPASNTAPQYIKANSKFIGITELTGYLYISAWPDSL 422
DB 358 HFKYCTAISGLDHLIPVAFKGDSTRTPTPLDPRELEILKTVKEITFTGLLIQAWPDN 417
QY 423 SVFONLOVIRGILHNGAYSITLQGLGTSWGLSLRELGLSLALIHNNHLCFVHTVPW 482
DB 418 HAFENLEIRKTKOHGOFSLAVVGLNITSLGLSLKEISDGDVVIISGNRLCYANTINW 477
QY 483 DQLFRNHQALLHTANRDEDECEVGLACHQLCARGHCWGPGPTQCVNCSQFLRQECVE 542
DB 478 KKLFGTNPQTKIMNNAEKDCKAVHVCNPLCSSEGCWGPEPRDCVSCQNVSRGECVE 537
QY 543 ECRVLQGLPREVYNARHCLPCHPEQCPQNGSVTCFGEADOCVACAHYKDPFFCVARCP 602
DB 538 KCNILEGPEPREFVENSECICQHPCEPLQPMNITCTGRGPDNCIOCAHYIDGPHCVKTC 597
QY 603 GVKPDLSTMPYWKPEDEGACQPCPINCTHSCVDLDDKGCFAEQASPLTSIVSAVVGIL 662
DB 598 GIMGENNTL-VMKYADANNVCHLCHANCTYGCAGFLQGCVEWFPSPKIPSIATGIVGGL 656
QY 663 LVVVLGVVFGI-LIKRROOKIRKVTMRRLLOETELVEPLTPSGAMPNOAMRILKETELR 721
DB 657 LFIVV-VALGICLPMRRRHIVKRTLRLLOERELVEPLTPSGEAPNOAHURLIKETEFK 715
QY 722 KVKVLGSAFTYVYKGIWIPGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGVGSFY 781
DB 716 KIKVLGSAFTYVYKGLMPEGEKVIKVAIKELREATSPKANKEILDEAYVMASVDNPH 775
QY 782 VSRLLGLCTSTVOLVTOLMYPYGLLDHVRNREGLSGODLLNMCWQIAKMSYLEVDRL 841
DB 776 VCRLLGLCTSTVOLITOLMYPYGLLDYVREHKONIGSQYLLANMCVQIAKMNYLEDRRL 835
QY 842 VHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMWALESIILRRFTH 901
DB 836 VHRDLAARNVLKTPQHVKITDFGLAKLLGAEKEYHAEGKVPKMWALESIILHRYTH 895
QY 902 QSDVWSYGVTVWELMTFGAKYDGIIPAREIPDLLEKGERLPPOPICTIDVYIMVVKCMI 961
DB 896 QSDVWSYGVTVWELMTFGSKYDGIIPASDISSILEKGERLPPOPICTIDVYIMVVKCMI 955
QY 962 DSECRPRFRELVSFSRMAPDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
DB 962 DSECRPRFRELVSFSRMAPDPQRFVVIQ-NEDLGPASPLDSTFYRSLLEDGMDGLVDA 1020
```

```
DB 956 DADSRPKFRELILEFSKWARDPQRYLVITQGDHMLPSPDTSNFPYRALMDEEDMEDVVDA 1015
QY 1021 BEYLVPQOGFFCPDAPGAGGMVHRRSSSTRSGGDLTGLGLEPSEEEAPRSLAPSEG 1080
DB 1016 DEYLTPOQGF-----NSPST-----SRTPLLSSLS 1041
QY 1081 AGSDVFDGLQMGAAKGLQSLPHTDPSPLORYSEDPTVPLPSET--DGYYAPLTCPOPE 1138
DB 1042 ATSN-----NSTVACINRNGSCRVKEDAFLOQYSSDPTGAVTEDNIDDAFL-----PVPE 1091
QY 1139 YVNPQDVTPQPPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAGFGGAVENPEYL- 1197
DB 1092 YVNO-SVPKRPAGSVQNPVYHNPQLHP-----APGRDLHYQN--PHSNAVGNPEYLN 1140
QY 1198 TPQGAAPQPPHPPAFSPAFDNLVYWDQ-----DP-----BERGAPSPSTFKGTP 1241
DB 1141 TAQ-----PTCLSSGFNSPALMIQKSHQMSLDNDPDYQODFFPKETRPNGIFKG-P 1190
QY 1242 TAENPEYLGLDVP 1254
DB 1191 TAENAEYLRVAPP 1203

RESULT 4
QYH40 PRELIMINARY; PRT; 1165 AA.
AC QYH40;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase proto-oncogene.
DS XMRK.
OS Xiphophorus xiphidium.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8086;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA MEDLINE=98241172; PubMed=9582016;
RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,
RA Altschmid J., Schartl M.;
RT "Activation of the Xmrk proto-oncogene of Xiphophorus by
RT overexpression and mutational alterations.";
RL Oncogene 16:1681-1690(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RIO PURIFICATION;
RA Schartl M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U53471; AAD10500.2; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CytC heme bind.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR01030; Recep_L domain; 2.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.
DR PROSITE; PS00343; GRAM_POS ANCHORING; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 2.
DR PROSITE; PS00103; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.
```

```
SQ SEQUENCE 1165 AA; 129614 MW; 7P7EE38D8771A74E CRC64;
Query Match 39.5%; Score 2692; DB 13; Length 1165;
Best Local Similarity 45.3%; Pred. No. 3.7e-194;
Matches 578; Conservative 163; Mismatches 387; Indels 148; Gaps 31;

QY 1 MELAALCRWGLLLALLPPG-AAST-----QVCTGDMKRLPASPETHLMRLHLYQCCQV 55
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4 LELEL-----LLLLLSIGRCCSTDPDRKVCQGTSMQTM---LDNHYLKKKMYSGCNV 56
QY 56 VQGNLELTYLPTNASLFLQDIOEVQGVLIHANVQVPLQRLRIVRGTOLFEDNYALA 115
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
57 VLENLEITYQENQDLSFLQSIQEVGYVLIANNEVSTIPLVNLRLIRGNLVEGNTLL 116
QY 116 VLNDGDLNNTTPTVGASPGGLRELQRLSRLTEILKGGVLIOENPOLCYQDILKWDIFHK 175
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
117 VMSNYQK-NPSSP--DYVQVGLQQLSLNTEILSGGVKSHNPLLCNVETINWMDIVDK 173
QY 176 NNQALTLIDTNRSRACHPCSPCKGSRCSGSESDCSLRTTRVCAGGC-ARCKGLPTD 234
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
174 TSNPTWNLIPIAFERQCKQDPCGVNGSCWAPGFGHCQKFTKLLCABQCNRRCRGPPID 233
QY 235 CCHQCAAGCTGPKGSHCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRTFGAS 294
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
234 CNEHCAGGCTGPRATDCLACROFNDGCTCKDTPPKIYDIIVSHQVVDNPNIKYTFGA 293
QY 295 CVTACPNYIUSTDVGSCSTLVCPLHNOEVTAEADGTQCEKSKPCARVCYGLM----QYI 350
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
294 CVKECPSNVYVTE-GACVRSASGMLQVD-ENGKRSCKPCDGVCPKVCDDIGIGLSLNTI 351
QY 351 KANSKEIGITELEFACKKIFGSLAPFSPESFDGDPASNTAPOVIKANSKEFIGITELTGYL 410
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
352 AVNSTWIG-----SFSNCTKINGDIILNRNSFEGDPHYKIGPMDPEHLNLTYYKEITGYL 407
QY 411 YISAWPDSLDLSVFQNLQVIRGILHNGAYS-LTLQGLGISMLGRLSLRELGSGLALIH 469
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
408 VIMWPNENITSLSVFQNLLEIRGRTTFSRCFSFVVQVSHLQWLGSLKEVSAGNVILK 467
QY 470 HNTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGACHOLCARGHCGPGTQCV 529
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
468 NTPQLRYASTINWRLFRSDQSIEDART-----ENQTCNECEDCGWPGPTMVCV 520
QY 530 NCSOFLRGQCEVECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAH 589
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
521 SCHVDGRGRCAVSNLLQGEPEA QVDRGCVQHQECLVQTDLSLTCYGFPGPANCCKCAH 590
QY 590 KYDPPFCVARCPGSGVXPDLSYMBIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPASORAS 649
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
581 FQDGPQCI PRCPHGMLGDGTL-IMKYADRWGQCQCHQNCQCGSGPGLSGCRGD-IVS 638
QY 650 PLTSIVSVAUGILLVVLGVVFGILIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQ 709
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
639 HSLAVGLVSGLLITVIVALLIIVLLRRRIK-RKRTIRLLQEKELVEPLTPSGQAPNQ 697
QY 710 AQWRILKETELRKVKVLGSGAGFTVYKGIWDGENVKIPVAIKVIRENTSPKANKEILD 769
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
698 AFRLIKETEFKADRVLGSGAGFTVYKGLNPNPGENIRIPVAIKVIREATSPKNOEVL 757
QY 770 EAYVMAGVSPVYRLLIGLCTSTVOLVTQLMYPYGLLDHVRNRRGLSGQDLLNWCMI 829
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
758 EAYVMASVDHPHVRLLIGLCTSAVOLVTQLMYPYGLLDYVRQHQRICQWLLNWCQVI 817
QY 830 AKGMSYLEDLVRLHRLAARNVLKSPNVHKITDFGLARLLDIDETEHADGCKVPIKWM 889
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
818 AKGMNYLEERHLVHRLAARNVLLKNPNHVKITDFGLSKLLTADKEYQAHGSKVPIKWM 877
QY 890 ALESILRRRFTHOSDWSVGVTVWELMTGAKPYDGI PAEIPDLLEKGERLPPOPICIT 949
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
878 ALESILQWITYTHOSDWSVGVTVWELMTFGSKPYDGI PAKETASVLENGERLPPOPICIT 937
QY 950 DVTMIMVKCMWIDSECRPRFRELVSFSEFRWARDPQRFFVITQNEIDLGPASPLDSTFVRSLL 1009
DB :|||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
938 EYVMILKCMWIDPSSRPFRFRELVGFEFSQWARDPFSYLVIOG-----NLPFSDRRLFSRL 994
```

## RESULT 5

```
Q9W6F6 PRELIMINARY; PRT; 1137 AA.
ID Q9W6F6; AC Q9W6F6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Receptor tyrosine kinase (Fragment).
GN ERB4.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HINDBRAIN;
RX MEDLINE=99263203; PubMed=10328884;
RA Dixon M., Lumsden A.;
RT "distribution of neuregulin-1 (nrg1) and erbB4 transcripts in
RT embryonic chick hindbrain.";
RL Mol. Cell. Neurosci. 13:237-258 (1999).
DR EMBL; AF121963; AAD31764.1; -.
DR HSSP; P11362; 1FGK
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001368; TNFR_c6.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR004019; YLP_motif.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00669; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR Pfam; PF02757; YLP; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00652; TNFR_NGFR_1; UNKNOWN_1.
KW Kinase; Tyrosine-protein kinase.
KW_NON_TER 1
SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;
```

```
Query Match 38.8%; Score 2646.5; DB 13; Length 1137;
Best Local Similarity 46.4%; Pred. No. 9.6e-191;
Matches 529; Conservative 173; Mismatches 359; Indels 79; Gaps 26;

QY 161 LCYQDTILWKDIFHKINQLALTIDTNRSRACHPCSPCKGSRCSGSESDCSLRTTVC 220
```

```
Db 3 LCFADTHWQDIVRNPWASFTLVPTNGSSGCRCHKSCGTG-RCWGPPTENHCQTLLTKTVC 61
QY 221 AGGC-ARCKGRLPTDCCHCCAACTGPKSDCLACLFHNSHGICELHCPALVTYNTDF 279
Db 62 AEQCDGRGCGYGVSDCHRECAAGSGPKDTCFACMNFNSGACVTCQPTFFYNPTTF 121
QY 280 ESMNPENGRYTFGASCVTACPYNYLSDVGSCTLVCPLNHOEVTAEDGTQRCCKSPCA 339
Db 122 QLEHNHNAKYTGAFCKVCKPHNFV-VDSSCVACPSRRKEV-EENGKWKCKPTDIP 179
QY 340 RVCVGLGMQYIKANSKFIGITELE-FAGCKIFGSLAPLPSPFGDPSASNTAPQYIKANS 398
Db 180 KACDGICTGSL-VSAQTVDSNIDKFNCTKINGNLIFLVTGIHGDPVHTIAALNPEKLN 238
QY 399 KFIGITELTGVLYLXSAMPDLSVFQNLQVIRGRILHNGAYSILTQGLGSLGRLSL 458
Db 239 IFQTVREITGVLTQSPENMTDFRFSNLVTIGRALYSLGLSLLILKQOQITSLQFQSL 298
QY 459 RELGSLALIHNTHLCFVHTVPWDQLFRNPHQALLHTANRDECEVGEGLACHQLCARG 518
Db 299 KQISAGNIYITNSNLCYHTVNTSLFSTPSQKTVIHRNKKAEKNTADGWVNCNELSSD 358
QY 519 HCWGPPTQCVNCSQFIRGQECVEECRVLOGLPREYVNAHCLPCHPEQCP-QNGSVTCF 577
Db 359 GCWGPDPQCLSKCFIRGRITCIESCNLYDGEFREFANGSVCMCEDPQCEKMDNMITCY 418
QY 578 GPEADQCVACAHYKDPFCVACRCSGVKPDLSYMPIMKFPDEEACQPCPINCTHSCVDL 637
Db 419 GPGPDHCTKCFHFGDGNCEKCPDGLQGANSF--IFKYADEDECHPCPNCTQGGCRGP 476
QY 638 DDGCG-----PAERASPLTSIVSAVV-GILLVVLGVVFGILIKRROKIRKY 685
Db 477 ASHDCIYVPTWQSTLPOHAR-TPL--IAAGVIGLFIIVIMGLTFAVYVRKSIK-KKR 532
QY 686 TMRLLEOTELVEPLTSGAMPNOAQMRILKETELRKVKULGSCAFGTGVKGIWPDGEN 745
Db 533 ALRRFL-ETELVEPLTSGTAPNAQALRIILKETELRKVKVLGSGAFGTGVKGIWVPEGT 591
QY 746 VKIPVAIKVLRENTSPRANKEILDEAYVMAGVGSPYVSRLLIGCLTSTVOLVTLQMPYGC 805
Db 592 VKIPVAIKILNETTGPRANVEFMDALIMASHDHPHVLRLGVCLSPITQLVTLQMPHGC 651
QY 806 LLDHVRNRRGLGSQDLLNMCQIAKMSYLEDLVRLVHRDLAARNVLVKSFNHVKIITDFG 865
Db 652 LLDVYHEHKDNIGSOLLNMCVQIAKGMVYLEERLVRDLAARNVLVKSFNHVKIITDFG 711
QY 866 LARLLDIDETEHADGSKVPIKMALESILRRRTHOSDVWSYGVTVWELMTFGAKPYDG 925
Db 712 LARLLEGEKEYNADGGKMPKMALECIHYRKFTHQSDVMSYGVTVWELMTFGKPYDG 771
QY 926 IPAREIPDLLEKGERLPPOPCTIDVYIMVWKMDSECRPRELVSSEFSRWARDPQR 985
Db 772 IPTREIPDLLEKGERLPPOPCTIDVYIMVWKMDADSRPKFELAAFSRWARDPQR 831
QY 986 FVVIQNEED-LGPAASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFCCPDPAAGAGGVH 1044
Db 832 YLVITQGDRLKPLSPNDSKFFQNLDEEDLEMDMAEYLVLP-QAFNIPPIYTSRTRID 890
QY 1045 HRRSSSTRSGGDLTLGLSPSEEAAPRS--PLAP-SEGAGSDVFDGLGMAAGKQLSL 1101
Db 891 SNRNQFVYRGGVAAEGV-PMPYRAGPCTIPEAPVAGATAEIFEEDTCNGTLRKQVAT 949
QY 1102 PTHDPSPLORYSEDPTVPLPS-----ETDGVYVAPLTCSPQPEYVNPQVRPOPSPRE 1154
Db 950 LAKEDSSTORYSADPTVFIPEVIRGELDEGDMTPMRDKPTDYLNPEENPVFSRRKN 1009
QY 1155 GPLPAA-RPAGATLERAKTUSPGKGVVQKVF-----AFGAVENPEYLTPOGGAAP 1205
Db 1010 GDLQAVDNPEYHN-----APNGQPKAEDEYVNEPLYLNTFANTLENAEYL----- 1054
QY 1206 QPHPPPPAFSFAFNLYWQDDQPERGA--PPSTFKGTPT-----AENPEYL 1249
```

```
Db 1055 KNNLPEKAKKAFNDPYWNHSLPPRSTLQHPDYLQLEYSTKYFYKQNGRIRPIVAENPEYL 1114
RESULT 6
P79754
ID P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE ErDB3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes.";
RL Genome Res. 9:251-258(1999).
DR HSSP; AF056116; AAC34391.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;
Query Match 33.2%; Score 2264; DB 13; Length 1328;
Best Local Similarity 40.0%; Pred. No. 9.3e-162;
Matches 515; Conservative 155; Mismatches 425; Indels 192; Gaps 32;
QY 9 WGLLLALLPP--GAASTQ-----VCTGDMKLRLPASPETHLDMLRHLYQGCVVQGNLEL 62
Db 4 WRLILMCVASRLRAASSQTQEA VCPGTQNGLSSTGSGENQYNLNKDRYKGCETIMGNLEI 63
QY 63 TYLPTNASLSFLQDIOEVQGVLIHNOVQVPLQRLRIVRGTLFEDNVALVLDNGDP 122
Db 64 TQESNWDFSLKTIREVTVGLVIAHMFQEIPLGQURVIRGNSLYERRFALSFLN--- 120
QY 123 LNNTTPVTGASPGGLRELQLRLSLTEILKGGVLIQORNPLCYQDTILMKDIFHKNNQLALT 182
Db 121 ----YPKDG--PSGLNQLGLMNLTEILDGGVQIINNKYLRYPVWVRDII--RNDAPIE 173
QY 183 LIDNRSRACHPCSPMKGRCWGESSEDQSLTRTVACGC-ARCKGRLPTDCCHCCAA 241
Db 174 IQFNGERGVCH---KSC-GNYCWPGRKQCCQIILTKTVCAPOCNDRCFGTSPRDCCHIECA 229
QY 242 AGCTGPKHSDCLACLFHNSHGICELHCPALVTYNTDFTFESMNPENGRYTFGASCVTACPY 301
Db 230 AGCKGRLDTCFACRLFNDSGACVPCQPTLIYNKQTFQMETNPNAKYQGSICVSCQPT 289
QY 302 NYLSTDVGSCTLVCPLNHOEVTAEDGTQR-CEKSKPCARVCYGLGMQYIKANSKFIGIT 360
Db 290 HFV-VDSSCVSVCPDPKMEV--ERGSQRCCLSGCLCPKVCETGAE----QRQTVDS 342
QY 361 ELE-FAGCKIKFGLSLAPLPSPFGDPSASNTAPQYIKANSKFIGITELTGVLYLXSAMP 419
Db 343 NIDSFINCTKIQSLHFLVLTGILGDDDFKQVPPPLDAKKLEVFRTVREITDILNIQSWPKEL 402
```

```
QY 420 PDLVSFQNLQVIRGRILHNGAYSLSLTQGLGSIWGLSLRSLRELGSGLALIHNNHLCVPHVT 479
DB 403 NDLSVFSSLTTCQGRSLFRFSVMVMRIPTLTSLGSLRSLREISDGSVYISQNAHLCLVHHT 462
QY 480 VPMQDLFRNPH-QALLHTANRDECEVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQ 538
DB 463 VNMVQLFRGSRVRANSNSRPAECVADGRVCDPLCSGCGWGPDPQCLSLNSRSHG 522
QY 539 ECVECRVLQGLPREYVNAH-CLPCHPECPQNGSVTCFPGADOCVACHYKDPDFCV 597
DB 523 TCVAGCHFNSGIPREFAGLNGVCVACHPECKPQTGRASCTGPGADECMACTKFRDGPYCM 582
QY 598 ARCPGSKVDPDLSYMPYKWPDEGACQPCPINCCHSCHVDLDDGCPAERASPLTSIVSA 657
DB 583 SSCPAGVN-DGEKGLIFKFPNREHCEPCQNCCTQCGSGPLNDC---LEAARLTSSGQ 638
QY 658 VVGILLVWLVGVF-----GILIKRQOKIRKYTMRRLLQBELTELVEPLTPSGAMPNOA 710
DB 639 ITGIALGVPAGLIFCLVFLGLMYHRLAIRKRAMRRYLESGESFEPLCP-GEKTKV 697
QY 711 QMRILKETELRKVKVLGSGAGFTYKGIWIPDGENKIPVAIKVLRENTSPKANKELDE 770
DB 698 HARILPSDLRKIKPLGSGVFGTVSGFWIPEGETVKIPVAIKTIQDSSGRQTFTEITDH 757
QY 771 AYVMAGVGSPIYSRLIGICLTSTVOLVTLQMPYGCILLDHVRENRLGSGQDILLNWCQIA 830
DB 758 LLSMGLSDHPYIVRLGICPGTCLQVLTQSSHSLLEHRIQHKTLDPQRLNWCQIA 817
QY 831 KGSYLEDVRLVHRLDAARNVLKSPNVKXITDFGLARLLDIDETEHADGKVPIKMA 890
DB 818 KGMVYLEHRVHVHKNLAARNLLKNDYQVQISDYGVADLLYPDKKYVYSETKTPKMA 877
QY 891 LESILRRRTHQSDVMVSGYVVMELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID 950
DB 878 LESILFRYTHQSDVMVSGYVVMWMSFGAEPYASVQBPVSVLEKGERLSQPAICTID 937
QY 951 VYIMVWKMWIDSECPREFELVSEFSRMARDQRFVVIQNEIDLGPASPLDSTFYSLLE 1010
DB 938 VYVMVWKMWIDENIRPTKELASDFTRWARDPRYLVRMEG-----E 981
QY 1011 DDMDGLVDAEEVYLVQGGFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1070
DB 982 DSGMGEFL-----RGRSER---GLLEADLEDEEE- 1008
QY 1071 PRSPLAPSEGASDVDFDGLGMG---AAKGLQLSLTHDPSPLQ-----RYSEDPT 1117
DB 1009 -----GLGDRFATPSLQSPSWSTSPSINSYMWVMTQLRYD----- 1044
QY 1118 VPLPSETDGVAPLTCSPQ-P-EYVQ-----PDVRPQPPSPREGPL--PAA 1160
DB 1045 --FAVSQGGHIGYLPMSPSVDITRQLWYQSRSLSSVRLTPDRSAFRSSRAELCEDGA 1102
QY 1161 RPAGATLERAKTILSPCKNGVVDVAFAGGAVENPEVLTPOGGAAPQHPPPAPFAPDNL 1220
DB 1103 QCAGIFRVR-----FGSERGN-----PQGG----- 1122
QY 1221 YVWDQPPPERGAPPSTFKTPTAENPE 1247
DB 1123 ---QQRKLSTASSPSSFKTWADEEDE 1146
```

## RESULT 7

```
Q9BIH9
ID Q9BIH9 PRELIMINARY; PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
```

```
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUA;
RA Lyceet G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000345; CytC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR P-coDom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Receptor.
FT NON_TER 1
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;
```

Query Match 29.5%; Score 2010.5; DB 5; Length 1433;  
Best Local Similarity 32.8%; Pred. No. 1.4e-142;  
Matches 471; Conservative 195; Mismatches 396; Indels 385; Gaps 38;

```
QY 26 CTGDMKRLPASPETHDMLRHLYQCCVQVQGNLELTYPNTNASLFLQDIOBQGVYL 85
DB 1 CIGTNGRMSVPANREHYKXNLRDRYNTCTYVDGNLEITWIONITDLNLFHQHIREVTGYL 60
QY 86 IAHNVQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTTPTVTGASPGGLREL 140
DB 61 ISLYDLQVILPLQIIRGRTTFKLNKWEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGGVLIQRNQPCYQDILWKDI-FHKXNQLALTLIDNRSRACHPCSPMC 199
DB 105 ELPALRDIILGSGVGFNNYNLCHKMSINWEEIILAPQTSMOYTFNFSSPERVCPCHPSC 164
QY 200 KGSRCWGESSEDCQSILTRTVACGGCA--RCKGPLPTDCHEQCAAGCTGPKHSDCLACLH 257
DB 165 EVG-CHWEGAHNCQRFSLKNCSPQCSQGRCFGPKPRECCHLFCAGGCTGTQSDCLACKN 223
QY 258 FNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYVLTSDVGSCVLCPL 317
DB 224 FYDDGVCKQECPPMQIYNPTNFWEPNPGKYATGATCVRKCP-EHLLKONGACVRKCPK 282
QY 318 HMQEVTAEQGTQCEKSKPCARVCYGLGMYIKANSKFGITELFEPAGCKKIFGSLAFL 377
DB 283 GKMPPQNSE-----CVPECKGVCPKTCFEGGI-----VHSDNIG-----NYKDCITLEGSLEIL 329
QY 378 PESFDGDPASNT-----APOYIKAN-----SKFIGITELTGYLYISAMPDLSLPDLSVFONLQ 429
DB 330 DQSGDFGQQVYTNFSGPRYIKIDPDRLEVFSTVKEITGFINQAHHPNTTLYNFANLE 389
QY 430 VIRGRILHNGAY-SLTLOGLIGISWLGRSLRELGSGLALIHNNHLCFVHTVPMQDLFRN 488
DB 390 VVGGRLKENLFASVYIVKTSLSLEKSLKRWNSGSIIVLENSDLCFDEIDWSEIKKS 449
QY 489 PHQALLHTANRDECEVGEGLACHQLCARGHCWGPPTQCVNCSQFLRGQECVEECRVLQ 548
DB 450 SDHEVMVQKNRNATECHEEGMECSQSKAGCWGKGPEQCLECKKNYKYGKCLDSCK--- 506
```





```
Query Match      25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 5.9e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESIILRRRTHOSDWSYGVTVWELMTFCAGKPYDGIIPAREIPDLLEKGERLPQPICT 948
Db 1 MALESIILRRRTHOSDWSYGVTVWELMTFCAGKPYDGIIPAREIPDLLEKGERLPQPICT 60

Qy 949 IDVTMIMVKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSL 1008
Db 61 IDVTMIMVKCMIDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGPASPLDSTFYRSL 120

Qy 1009 LEDDMGDLVDAEYLVPQGFCDPAPGAGGMVHRHRSSSTRSGGDLTLGLEPSEE 1068
Db 121 LEDDMGDLVDAEYLVPQGFCDPAPGAGGMVHRHRSSSTRSGGDLTLGLEPSEE 180

Qy 1069 EAPRSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 1128
Db 181 EAPRSLAPSEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYV 240

Qy 1129 APLTCSQPEYVNOQVVRPPSPREGPLPAARPAGATLERAKTLSPGKGVKDVFAFG 1188
Db 241 APLTCSQPEYVNOQVVRPPSPREGPLPAARPAGATLERAKTLSPGKGVKDVFAFG 300

Qy 1189 GAVENPEYLTPOGAAQPHPPRAFSFADNLYYWDQDPPERGAPPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLTPOGAAQPHPPRAFSFADNLYYWDQDPPERGAPPSTFKGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 ID Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE POLYPROTEIN.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retrov. viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSSP; P03322; 1A6S.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF02813; Retro.M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 842F6914EFE1D63 CRC64;

Query Match      25.2%; Score 1720; DB 15; Length 729;
Best Local Similarity 84.8%; Pred. No. 4.5e-121;
Matches 358; Conservative 75; Mismatches 118; Indels 102; Gaps 15;
```

```
Qy 569 PQSGSVTCFGEADQCACAHYKDPFPCVACRPSGVKPDLSYMPPIWKFDPDEGACQPCPI 628
Db 141 PEETATPKTGP--DHCMAKAFIDGPHCVACRPAAGVLGENDTL--VMKYADANAVCOLCHP 197

Qy 629 NCHTSCVDLDDKCPAQRASPLTISVAVV--GILLVVLGVVFGILIKRQOKIRKYTM 687
Db 198 NCTRGCKGPLEGCP---NGSKTFSIAAGVVGGLCLVVGVLGILGLYURRR--HIVKRRTL 253

Qy 688 RRLLOETELVEPLTPSGAMPNQAMRILKETELRKVKVLGSGAGFTVYKGIWIPDGVNK 747
Db 254 RRLLOERELVEPLTPSGEAPNQAHILKETEFKVKVVLGSGAGFTVYKGIWIPGEKVK 313

Qy 748 IPVAIKVLRENTSIPKANKEILDEAYVMAGVSPVYSRLGLICTTSTVOLVTQLMPYGCCL 807
Db 314 IPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGICLTSTVQLITQLMPYGCCL 373

Qy 808 DHVRENFRGLSGODLNNWQIAKMSYLEDVRLVHRDLAARNVLVKSNNHVKITDFGLA 867
Db 374 DYIREHKDNTGSGYLLNWCVOIAKGMNLYBEERLVLVHRDLAARNVLVKTPOHVKITDFGLA 433

Qy 868 RLIDIDETEHADGCKVPIKMMALESILRRRFTHQSDVMSYGVTVWELMTFCAGKPYDGIP 927
Db 434 KLLGADKEVHAEGCKVPIKMMALESILHRIYHQSDVMSYGVTVWELMTFCGSPYDGIP 493

Qy 928 AREIPDLLEKGERLPQPICTIDVYIMVKCMIDSECRPRFRELVSFSESRMARDPQRFV 987
Db 494 ASEISSVLEKGERLPQPICTIDVYIMVKCMIDADSRPKFRLEIAEFSSKWARDPPRYL 553

Qy 988 VIQ-NEDLGPASPLDSTFYRSLLEDDMGDLVDAEYLVPQGFCDPAPGAGGMVHR 1046
Db 554 VIQGERMHLPSPDTSKFYRLMBEEDMEDIVDAEYLVPHQGF----- 598

Qy 1047 HRSSSTRSGGDLTLGLEPSEEAPRSL-----APSEGAGSDVFDGDLGMAAGLQSL 1101
Db 599 -NSPST-----SRTPLLSSLSATSNNSATNCID-----RNGOGH 631

Qy 1102 PTHDPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQPEYVNOQVVRPPSPREGPLPA 1159
Db 632 PVREDSFVQRYSSDPTGNFLEESIDDGFL-----PAPEYVNO--LMPKPS----- 675

Qy 1160 ARPAGATLERAKTLSPGKGVKDVFP-----AFGGAIVENPEYL 1197
Db 676 -----TAMVQNOIYNNISLTAISKLPMDSRYNQNSHSTAVDNPEYL 715

RESULT 11
Q86714 ID Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retrov. viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
```



```
DR EMBL; X52211; CAA36459.1; JOINED.
DR HSSP; P10828; 2NLL.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR001245; Tyr_pkinase.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000001; Euk_pkinase; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00219; TyrKc; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW Transcription regulation; Transferase; Tyrosine-protein kinase;
KW Zinc-finger.
SQ SEQUENCE 962 AA; 108320 MM; 3C5AED791E4E95CE CRC64;

Query Match 24.4%; Score 1662.5; DB 15; Length 962;
Best Local Similarity 39.7%; Pred. No. 1.5e-116;
Matches 420; Conservative 115; Mismatches 253; Indels 271; Gaps 37;

QY 248 KHSCLAC-----LHNHSGICELHCPALVTYNTDTFESMPNEGRTYTCASCVTACPY 301
DB 65 KDEQCVCGDKPTGYHYR-----C-ITCEGCKSFRRTIQKNLHP-----TYSCTY 109

QY 302 NYLSTDVGSCTLVCLPHNQEVTAEDGTQCEKCS-KPCARVCYGLGMQYIKANSKFTGIT 360
DB 110 D-----GCCVIDKITRNQ-----CQLCRFKKICISV-GWAMDVLVDDSKRAKR 151

QY 361 EL-----EPAGCKKIFGSLAFPS-----PDGPASNTAPQYIKANSKFI--- 401
DB 152 KLTEENRRRRKEEMIKSLQHRPSPSAEWEELIHVVTEAHRSTNAQGSWHKQRKFLLED 211

QY 402 -----GITELTGYL--VISAWPDSLPLDSVFQNLQVIRGRILLNG-- 439
DB 212 IGQSPHASMLDGDKVDLEATEFTKIIITPAITRVVDFAKNLPMSFELPCEDQIILLGCC 271

QY 440 -----AYSILTQG-----LGISWTLGRS--LRELGSGLALIH-HNTH 473
DB 272 MEIMSLRAAVRYDPESETLTLSGEMAVKREQLKNGGLGVVSDAIFDLGKLSAFNLDDE 331

QY 474 LCFVHTVPWDLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPGTQVCNCSQ 533
DB 332 VALLOAV-----LLMSDR-----TGLIC----- 350

QY 534 FLRQGEQCEVERCVLQGLPRE-YVNAR-HCLP----- 562
DB 351 -----VDKIEKQCSYLLAFPHYINRYKHNIPHWFSKLLMKVADLRMTGAVHARFLHMV 406

QY 563 -CHPECOQNGSVTCFQPEADQCACAHYKDDPFCVACRPSGVKPDLSYMPIWKFPDDEG 621
DB 407 ECPTELSPOE-----VGP--DHCWKCAHFIDGPHCVKACPCAGVLGENDTL-VMKYADANA 458

QY 622 ACQPCPINTHSCVDLDDKCPAEQASPLTISVAVV-GILLVVLGVVFGILIKRQ 680
DB 459 VCQLCHPNCTRGCKPGLEGCP---NGSKTPSIAAGVVGGLCLLVVVGILGILYLR--H 514

QY 681 KIRKYMRLRLQELTELVEPLTPSGAMPNQOMILKETURKVKVILGSGAGFTVYGIWI 740
DB 515 IVRKRTLRLRLQRELVEPLTPSGEAPNQAHILKETEFKVKVILGFGAGFTVYKGLWI 574

QY 741 PGENYKIPVAIKVLRNTSPKANKEILDEAYVMAGVSPYVSRLLGICLTSTVQLVTLQ 800
DB 575 PEGEKVTIPVAIKELREATSPKANKEILDEAYVMASVDNPHVCRLLGICLTSTVQLITQL 634
```

```
QY 801 MPYCGLLDHVRENRLGSGQDLLNWCQIAKMSYLEVDRLVHRDLAAARNVLVSPNHVK 860
DB 635 MPYCGLLDYIREHKDNIGSQYLLNWCQIAKGMVYLBERRHVMVHRDLAARNVLVKTPOHVK 694

QY 861 ITDEGLARLLDIDETEYHADGGKVPKWKMALESILRRRFTHQSDVWSYGVTVWELMTGA 920
DB 695 ITDFGLAKQLGADEKEYHAEGGKVPKWKMALESILHRIYTHQSDVWSYGVTVWELMTFGS 754

QY 921 KPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVWKCMWLDSECRPRFRELVSFSRMA 980
DB 755 KPYDGIIPASEISSVLEKGERLPQPPICTIDVYIMVWKCMGSGARPRFRELIAFSGMA 814

QY 981 RDPQRFVVIQ-NEDLGPAASPLDSTFYRSLLDDDDMDGLVDAAEYLVPOQGFCCPDPA 1039
DB 815 RDPPEYLVICQDERMHLPSPTDSKFYRTLMBEEDMEDIVDAEYLVPHQGFF----- 866

QY 1040 GGMVHHRHRSSTRSGGDLTLGLEPSEEAAPRSPFLAPSEGAGSDVDFCDLGMGAAGLQ 1099
DB 867 -----NSPST-----SRTPLLSLSATSN-----NSATKCID 893

QY 1100 SLPTHDPSPQRYSEDTVPLPSETDGVVAPLTCSPOPEYVNPQDVRPQPPSPREGPLPA 1159
DB 894 RNCGH-----PVREDGFL-----PAPEYVQ--LMPKPFSTAMVQNOI 929

QY 1160 ARPAGAT-LERAKTLPSPKNGVWKDVFAFGGAVENPEYL 1197
DB 930 YNYISLTAISKLPMSRYQN-----SHSTAVDNPEYL 961

RESULT 14
QY 85468 PRELIMINARY; PRT; 545 AA.
AC Q85468;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Avian Erythroblastosis virus (Te34) v-erbB gene.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88217326; PubMed=2897102;
RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;
RT "Common site of mutation in the erbB gene of avian erythroblastosis
RT virus mutants that are temperature sensitive for transformation.";
RL Oncogene Res. 1:265-278 (1987).
DR EMBL; X06943; CAA30024.1; -.
DR HSSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 545 AA; 68899 MM; 140DCE8CCA0F8AF4 CRC64;

Query Match 24.1%; Score 1645; DB 15; Length 545;
Best Local Similarity 54.9%; Pred. No. 1.3e-115;
Matches 345; Conservative 70; Mismatches 121; Indels 92; Gaps 15;

QY 578 GPADQCVACAHYKDDPFCVACRPSGVKPDLSYMPIWKFPDDEGACQPCPINTHSCVDL 637
DB 1 GP--DHCWKCAHFIDGPHCVKACPCAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGP 57

QY 638 DDGKCPAEQASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKYTWRLRLQTEL 696
DB 58 GLEGCP---NGSKTPSIAAGVVGGLCLLVVVGILGILYLR--HIVKRTLRLRLQREL 113
```

```
Qy 697 VEPLTPSGAMPNOAQRILKETELRKVKVLGSGAGFYVYKGINWIPDGENVKIPVAIKVL 756
D 114 VEPLTPSGEAPNOAHLRLKETEPKVKVLGFGAGFYVYKGLWIPEGEKVTPVAIKELR 173
Qy 757 ENTSPKANKEILDYAYWAGVGSYVSRLLGICLTSTVOLVQLMPYGCCLLDHVRNGR 816
D 174 EATSPKANKEILDYAYWASVDNPHVCRLLGICLTSTVOLVQLMPYGCCLLDYIREHKN 233
Qy 817 LGSQDLNWCWQIAKMSYLEDVRLDRAARNVLKSPNHVKITDPLGLARLLDIDETE 876
D 234 IGSQYLLNWCQIAKGNLYLEERHLVHRDLAARNVLKTPQDVKITDPLGLAKOLGADEKE 293
Qy 877 YHADGGKVPKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGPAREIPDLLE 936
D 294 YHAEKGKVPKMALESILHRIYTHQSDVWSYGVTVWELMTFGSKPYDGPASEISSVLE 353
Qy 937 KGERLPOPPCTIDVYIMVKCMWIDSECRPRFRELVESESRMARDPQRFVVIQ-NEDLG 995
D 354 KGERLPOPPCTIDVYIMVKCMWSDADSRPKFRELIAEFSKMARDPPRYLVIQGDERMH 413
Qy 996 PASPLDSTFYRSLLEDMDGLVDAAEYLVPOQGFPCPDAPAGAGGMVHHRRSSSTRSG 1055
D 414 LPSPDTSKFYTLNEEDMEDIVDAEYLVPHQGF-----NSPST--- 454
Qy 1056 GGDLTGLPESEEAAPSPL-----APSEGAGSDVFDGDLGMGAAGLQSLPTHDPSPLO 1110
D 455 -----SRTPLLSSLSATSNNSATNCIDRNGG-----H----- 481
Qy 1111 RYSEDTPVLPSETDGVYAPLTCSPQBYVYNQDVRPQPSPREGPIPAARAGAT-LER 1169
D 482 -----PVREDGFL-----PAPEYVNO--LMPKPKSTAMVQIQIYNYISLTISK 523
Qy 1170 AKTLPSPKNGVVKDVFAGGAVENPEYL 1197
D 524 LPWDSRYQN-----SHSTAVDNPEYL 544

RESULT 15
Q9MVFS PRELIMINARY; PRT; 655 AA.
AC Q9MVFS;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L.; Threadgill D.W.; Danielson A.J.; Schell C.;
RA Lampland A.L.; Balasubramaniam S.; Crossley T.O.; Magnuson T.R.;
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV7TAC;
RA Reiter J.L.; Threadgill D.W.; Eley G.D.; Strunk K.E.; Danielson A.J.;
RA Sinclair C.S.; Pearshall R.S.; Green P.J.; Yee D.; Lampland A.L.;
RA Balasubramaniam S.; Crossley T.O.; Magnuson T.R.; James C.D.;
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J.; Shinagawa A.; Shibata K.; Yoshino M.; Itoh M.; Iehii Y.;
RA Arakawa T.; Hara A.; Fukunishi Y.; Konno H.; Adachi J.; Fukuda S.;
RA Azawa K.; Izawa M.; Nishi K.; Kiyosawa H.; Kondo S.; Yamanaka I.;
RA Saito T.; Okazaki Y.; Gojobori T.; Bono H.; Kasukawa T.; Saito R.;
RA Kadota K.; Matsuda H.A.; Ashburner M.; Batalov S.; Casavant T.;
RA Fleischmann W.; Gaasterland T.; Gissi C.; King B.; Kochiwa H.;
RA Kuehl P.; Lewis S.; Matsuo Y.; Nikaide I.; Pesole G.; Quackenbush J.;
RA Schriml L.M.; Staubli F.; Suzuki R.; Tomita M.; Wagner L.; Washio T.;
RA Sakai K.; Okido T.; Furuno M.; Aono H.; Baldarelli R.; Barch G.;
RA Blake J.; Boffelli D.; Bojunga N.; Carninci P.; de Bonaldo M.F.;
RA Brownstein M.J.; Bult C.; Fletcher C.; Fujita M.; Gariboldi M.;
RA Gustincich S.; Hill D.; Hofmann M.; Hume D.A.; Kamiya M.; Lee N.H.;
RA Lyons P.; Marchionni L.; Mashima J.; Mazzarelli J.; Mombaerts P.;
RA Nordone P.; Ring B.; Ringwald M.; Rodriguez I.; Sakamoto N.;
RA Sasaki H.; Sato K.; Schoenbach C.; Seya T.; Shibata Y.; Storch K.-F.;
RA Suzuki H.; Toyooka K.; Wang K.H.; Weitz C.; Whittaker C.; Wilming L.;
RA Wynshaw-Boris A.; Yoshida K.; Hasegawa Y.; Kawai H.; Kohtsuki S.;
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:585-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGD; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR SMART; SM00261; Fu; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 21.4%; Score 1458.5; DB 11; Length 655;
Best Local Similarity 43.8%; Pred. No. 2.1e-101;
Matches 280; Conservative 94; Mismatches 241; Indels 25; Gaps 9;

Qy 11 LLLALLPPGAA--STQVCTGTDMLKLPASPETHLMLRHLQYQCVVQGNLEITYLPTN 68
D 14 LTTALCAAGALEEKVKVCGQTSNRLTQLGTGFEDHFLSLQRMNCEVVLGNLEITYQRN 73
Qy 69 ASLSFLDIOIEVOGYLIAHNQVRQVPLQRLIVRGTLFEDNYVALAVLQNGDPLNNTTP 128
D 74 YDLSFLKTIQEVAGYVLIALTVERIPLNQLIIRGNALYENTYVALAISL----- 124
Qy 129 VTGASPGLELOLRSLTEILKGVLIQRNPOLCYQDTILWKDI----FHKNQALTLI 184
D 125 -YGNRTGLRELPMRNLLQELIGAVRFSNNPILCNMDTIQWRDIVQNVFMSNMSMDL--- 180
Qy 185 DTRNRACHPCSPWCKGRCWGSESDCQSLTRTVACAGCA-RCKGPLPTDCHEQCAAG 243
D 181 -QSHPSCKPKDCPCNGSCWGGEENCKLTKIQAQCSHRCGRSPDCCHNQCAAG 239
Qy 244 CTGPKHSDCLACLHFNHSGICELHCPALVYNTDTTFESMPNPEGRYTFGASCVTACPNY 303
D 240 CTGPRESDCILVCQKFODEATCKDTPPLMLNPTTYQMDVNVPEGKSFATCVKCKPRNY 299
Qy 304 LSTDVGSCTLVCPLNHQEVTAEGTORCEKSPCARVCYGLGMYIKANSKFGITELE 363
D 300 VTDHSGSVACGPDYVEV-EEDGIRKCKCDGPRKVCNGIGIGEFK-DTSLINATNK 357
Qy 364 -FAGCKKIFGLSFLAPESPFDGDPASNTAPOYIKANSKFIGITELTGLVYTSAPWDSPLD 422
D 358 HFKYCTAISGLHLPLVAFKGDSTRTPLDPRLEILKTVKEITGELLQAMPDNDWTL 417
Qy 423 SVFQNLQVIRGRILHNGAVSLTQGLGISWGLRSLRELGSGLALIHNTLHLCFVHTVPW 482
D 418 HAFENLEIIRGTRTKHQHGFSLAVVGLNITSLGRSLKEISDGDVVISGNEINLCVANTINW 477

SEQUENCE
```

Qy	483	DQLFRNPQALLHTANPEDECVGEGLACHQLCARGHCWGPGPTQCVNCSQFLRGQECVE	542
Db	478	KKLFGTENQTKIMNNRAEKDKAVNHVCNPLCSSEGCWGPEPRDCVSCQNVSRGECVE	537
Qy	543	ECRVLOGLPREYVNARHCLFCHPECPONGSVTCFGEADOCVACAHYKDPFCVACPS	602
Db	538	KCNILEGEPREFVENSECIOCHPECLFOAMNITCTGRGPDNCIOCAHYIDGPHCVKTCPA	597
Qy	603	GVKPDLSYMPITWKFPEDEGACOPCPINCTHSCVDLDDKGC	642
Db	598	GINCENTL-VWKYADANNVCHLCHANCTYCCAGPGLOGC	636

Search completed: July 22, 2003, 09:01:11  
Job time : 53.3575 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds  
(without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-369-383-12

Perfect score: 6815  
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVFPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6653	97.6	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5865	86.1	1257	1 ERB2_RAT	P06494 rattus norv
3	5851.5	85.9	1254	1 ERB2_MESAU	Q60553 mesocricetu
4	3095	45.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3070	45.0	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2919.5	42.8	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2900	42.6	1308	1 ERB4_RAT	Q62956 rattus norv
8	2652.5	38.9	1167	1 XMRK_XIPWA	P13388 xiphophorus
9	2375.5	34.9	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2303.5	33.8	1339	1 ERB3_RAT	Q62799 rattus norv
11	1922	28.2	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB_ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB_AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB_AVIEU	P11273 avian eryth
15	1553	22.8	703	1 EGFR_CHICK	P13387 gallus gall
16	1275	18.7	1323	1 LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculu
18	730	10.7	1363	1 ILPR_BRALA	O02466 branchiosco
19	696	10.2	1383	1 INSR_RAT	P15127 rattus norv
20	696	10.2	1607	1 MIPR_LYMST	Q25410 lymnaea sca
21	695.5	10.2	1372	1 INSR_MOUSE	P15208 mus musculu
22	695	10.2	1382	1 INSR_HUMAN	P06213 homo sapien
23	689	10.1	1300	1 IRR_MOUSE	Q9WCL4 mus musculu
24	682	10.0	1297	1 IRR_HUMAN	P14616 homo sapien
25	676	9.9	1477	1 HTK7_HYDAT	Q25197 hydra atten
26	674.5	9.9	1300	1 IRR_CAVPO	P14617 cavia porce
27	644	9.4	1367	1 IGR_HUMAN	P08069 homo sapien
28	627	9.2	1390	1 INSR_AEDAE	Q93105 aedes aegypt
29	626	9.2	1373	1 IGR_MOUSE	Q60751 mus musculu
30	622.5	9.1	1370	1 IGR_RAT	P24062 rattus norv
31	612	9.0	2146	1 INSR_DROME	P09208 drosophila
32	599	8.8	987	1 EPB4_HUMAN	P54760 homo sapien
33	588	8.6	1114	1 RET_HUMAN	P07949 homo sapien

#### RESULT 1

##### ERB2\_HUMAN

ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erbB-2 gene to epidermal growth factor receptor.";  
RL Nature 319:230-234 (1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Cousseins L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139 (1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization.";  
RL Genomics 15:426-429 (1993)  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

#### ALIGNMENTS

34	587.5	8.6	984	1	EPB1_RAT	P09759 rattus norv
35	585.5	8.6	984	1	EPB1_CHICK	Q07494 gallus gall
36	581.5	8.5	977	1	EPB2_MOUSE	Q03145 mus musculu
37	581.5	8.5	984	1	EPB1_HUMAN	P54762 homo sapien
38	579	8.5	976	1	EPB2_HUMAN	P29317 homo sapien
39	576	8.5	902	1	EPBB_XENLA	Q91736 xenopus lae
40	573.5	8.4	1053	1	FAK1_CHICK	Q00944 gallus gall
41	570	8.4	987	1	EPB4_MOUSE	Q91571 xenopus lae
42	569.5	8.4	985	1	EPBA_XENLA	P53356 hydra atten
43	569	8.3	757	1	HT16_HYDAT	Q91738 xenopus lae
44	569	8.3	1068	1	FAK1_XENLA	P34152 mus musculu
45	563	8.3	1052	1	FAK1_MOUSE	

CC CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
CC CC RESIDUES (BY SIMILARITY).  
CC CC -!- POLYMORPHISM: THERE ARE FOURS ALLELES DUE TO THE VARIATIONS IN  
CC CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
CC CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
CC CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
CC CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC CC -----  
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC CC the European Bioinformatics Institute. There are no restrictions on its  
CC CC use by non-profit institutions as long as its content is in no way  
CC CC modified and this statement is not removed. Usage by and for commercial  
CC CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC CC -----  
DR EMBL; M11767; AAA35808.1; -  
DR EMBL; M11761; AAA35808.1; JOINED.  
DR EMBL; M11762; AAA35808.1; JOINED.  
DR EMBL; M11763; AAA35808.1; JOINED.  
DR EMBL; M11764; AAA35808.1; JOINED.  
DR EMBL; M11765; AAA35808.1; JOINED.  
DR EMBL; M11766; AAA35808.1; JOINED.  
DR EMBL; M11730; AAA75493.1; -  
DR EMBL; M12036; AAA35978.1; -  
DR EMBL; X03363; CAA27060.1; -  
DR PIR; A25491; A25491.  
DR PIR; A24571; A24571.  
DR HSSP; P11362; 1FGK.  
DR GenBank; HGNC:3430; ERBB2.  
DR MIM; 164870; -  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TyrK; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Polymorphism.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 653 675 POTENTIAL.  
FT DOMAIN 676 1355 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 220 227 BY SIMILARITY.  
FT DISULFID 224 235 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.

FT	DISULFID	563	576	BY SIMILARITY.
FT	DISULFID	567	584	BY SIMILARITY.
FT	DISULFID	587	596	BY SIMILARITY.
FT	DISULFID	600	623	BY SIMILARITY.
FT	DISULFID	626	634	BY SIMILARITY.
FT	DISULFID	630	642	BY SIMILARITY.
FT	MOD_RES	1139	1139	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	1248	1248	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOHYD	68	68	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	124	124	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	187	187	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	259	259	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	530	530	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	571	571	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD	629	629	N-LINKED (GLCNAC...) (POTENTIAL).
FT	VARIANT	654	654	I -> V.
FT	VARIANT	655	655	/FTid=VAR_004077.
FT	VARIANT	655	655	I -> V.
FT	CONFLICT	1170	1170	/FTid=VAR_004078.
FT	CONFLICT	1170	1170	P -> A (IN REF. 2).
SQ	SEQUENCE	1255	AA; 137909	MM; 39E9DFDA04DCF962 CRC64;

Query Match 97.6%; Score 6653; DB 1; Length 1255;  
Best Local Similarity 97.6%; Pred. No. 0;  
Matches 1225; Conservative 8; Mismatches 22; Indels 0; Gaps 0;

QY	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTDMLRLPAS	PETHLDMLRHL	YQCCVQGNL	60	
DB	1	MELAAALCRWGLLLALLPPGAAS	TQVCTGTDMLRLPAS	PETHLDMLRHL	YQCCVQGNL	60	
QY	61	ELTYLPTNASLSFLQDIOEVQGYVLI	IAHQVRQVPLQRL	LRVGRGTLFED	NYALAVLNG	120	
DB	61	ELTYLPTNASLSFLQDIOEVQGYVLI	IAHQVRQVPLQRL	LRVGRGTLFED	NYALAVLNG	120	
QY	121	DPLNNTTPTVGASPGGLRELOLRSL	TEILKGGVLIQRP	QCYQDTILWKD	IFHKNNQLA	180	
DB	121	DPLNNTTPTVGASPGGLRELOLRSL	TEILKGGVLIQRP	QCYQDTILWKD	IFHKNNQLA	180	
QY	181	LTLIDTNRSRACHPCSPKRCWGSE	SDCQSLTRTV	CAGGCARCKGL	PTDCCHEOC	240	
DB	181	LTLIDTNRSRACHPCSPKRCWGSE	SDCQSLTRTV	CAGGCARCKGL	PTDCCHEOC	240	
QY	241	AAGCTGPKHSDCLACLFHNSGIC	ELHCPALVYNT	DTFESMPNPEGR	YTFGASCVTACP	300	
DB	241	AAGCTGPKHSDCLACLFHNSGIC	ELHCPALVYNT	DTFESMPNPEGR	YTFGASCVTACP	300	
QY	301	YNYLSTDVGSCTLVCP	PLHNQEVTAEDGTQ	RCEKSKPCARVCY	GLGMOYIKANSK	FIGIT 360	
DB	301	YNYLSTDVGSCTLVCP	PLHNQEVTAEDGTQ	RCEKSKPCARVCY	GLGMOYIKANSK	FIGIT 360	
QY	361	ELEFAGCKKIFGSLAF	LPESFDGDPASNTA	POYIKANSKFCI	TELTGVL	YISAMPDSL 420	
DB	361	IOEFAGCKKI	FGSLAF	PESFDGDPASNTA	LPQEPQLV	ETLEITGVL	YISAMPDSL 420
QY	421	DLVSFQNLQVIRGRILHNGAY	SLTLQGLIGISW	LGRLSRLREL	SGSLALIH	NTHLCFVHTV 480	
DB	421	DLVSFQNLQVIRGRILHNGAY	SLTLQGLIGISW	LGRLSRLREL	SGSLALIH	NTHLCFVHTV 480	
QY	481	PWDQLFRNPHOALLHTANR	PEDECVCEGLACH	QICARGHCWG	PGPTQCNC	SQFLRGQEC 540	
DB	481	PWDQLFRNPHOALLHTANR	PEDECVCEGLACH	QICARGHCWG	PGPTQCNC	SQFLRGQEC 540	
QY	541	VEECRVQLGLPREVYNARHCL	PCHEPCOPNGSV	TCFGEADQC	VACAHYKDP	PPFCVARC 600	
DB	541	VEECRVQLGLPREVYNARHCL	PCHEPCOPNGSV	TCFGEADQC	VACAHYKDP	PPFCVARC 600	
QY	601	PSGVKPDLSYMPIWKFP	DEEGACQPCPINC	THSCVDLDD	KGCPAEQ	RASPLTSI	VSAVVG 660
DB	601	PSGVKPDLSYMPIWKFP	DEEGACQPCPINC	THSCVDLDD	KGCPAEQ	RASPLTSI	VSAVVG 660
QY	661	ILLVVVLGVVFGIILIKRRQ	KIRKYTWRRLLQ	ETELVEPL	TPSGAMP	NAQMRIL	KETEL 720
DB	661	ILLVVVLGVVFGIILIKRRQ	KIRKYTWRRLLQ	ETELVEPL	TPSGAMP	NAQMRIL	KETEL 720



Qy 721 RKVVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780  
 Db 721 RKVVLGSGAGTGYKGIWIPDGENVKIPVAIKVLRENTSPKANKEILDEAYVMAGVGP 780  
 Qy 781 YVSRLLGLICTSTVOLTPYGLDVRNENGRGLSQDLLNWCMIAGKMSYLEVDV 840  
 Db 781 YVSRLLGLICTSTVOLTPYGLDVRNENGRGLSQDLLNWCMIAGKMSYLEVDV 840  
 Qy 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
 Db 841 LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900  
 Qy 901 HQSDVWSVGVTVWELMTFGAKPYDGPAREIPDILEKGERLPPOPICTIDVYIMVKWM 960  
 Db 901 HQSDVWSVGVTVWELMTFGAKPYDGPAREIPDILEKGERLPPOPICTIDVYIMVKWM 960  
 Qy 961 IDSECRPRFRELVSFEFSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMDGLVDA 1020  
 Db 961 IDSECRPRFRELVSFEFSMARDPQRFVVIQNEGLGPASPLDSTFYRSLLDDMDGLVDA 1020  
 Qy 1021 EYILVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
 Db 1021 EYILVPOQGFPCPDAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
 Qy 1081 AGSDVDFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBYV 1140  
 Db 1081 AGSDVDFDGLGMAKGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQBYV 1140  
 Qy 1141 NOPDVRPOPSPREGPIPAARPAAGATLERAKTLSPGKGVVDVFAFGGAVENPEYLTPO 1200  
 Db 1141 NOPDVRPOPSPREGPIPAARPAAGATLERAKTLSPGKGVVDVFAFGGAVENPEYLTPO 1200  
 Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVFPV 1255  
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVFPV 1255

RESULT 2  
 ERB2 RAT STANDARD; PRT; 1257 AA.  
 AC P06494;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor  
 DE receptor-related protein).  
 GN ERB2 OR NEU.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RX MEDLINE=86118662; PubMed=3945311;  
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;  
 RT "The neu oncogene encodes an epidermal growth factor receptor-related  
 RT protein.";  
 RL Nature 319:226-230 (1986).  
 RN (2)  
 RP SEQUENCE OF 852-905 FROM N.A.  
 RC TISSUE=Sciatic nerve;  
 RX MEDLINE=91222560; PubMed=2025425;  
 RA Lai C., Lemke G.;  
 RT "An extended family of protein-tyrosine kinase genes differentially  
 RT expressed in the vertebrate nervous system.";  
 RL Neuron 6:691-704 (1991).  
 RN (3)  
 RP STRUCTURE BY NMR OF 650-668.  
 RX MEDLINE=92155181; PubMed=1346763;  
 RA Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,

RA Newman R., Crumpton M.J., Sternberg M.J.B., Campbell I.D.;  
 RT "Three dimensional structure of the transmembrane region of the proto-  
 RT oncogenic and oncogenic forms of the neu protein.";  
 RL EMBO J. 11:43-48 (1992).  
 CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP10 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.  
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X03362; CAA27059.1; ALT\_INIT.  
 DR PIR; A24562; TVRTNU.  
 DR HSP; P11362; IFGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR004019; YLP motif.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 655 677 POTENTIAL.  
 FT DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 159 369 CYS-RICH.  
 FT DOMAIN 473 646 CYS-RICH.  
 FT DOMAIN 722 989 PROTEIN KINASE.  
 FT NP\_BIND 728 736 ATP (BY SIMILARITY).  
 FT BINDING 755 755 ATP (BY SIMILARITY).  
 FT ACT\_SITE 847 847 BY SIMILARITY.  
 FT DISULFID 196 205 BY SIMILARITY.  
 FT DISULFID 200 213 BY SIMILARITY.  
 FT DISULFID 221 228 BY SIMILARITY.  
 FT DISULFID 225 236 BY SIMILARITY.  
 FT DISULFID 237 245 BY SIMILARITY.  
 FT DISULFID 241 253 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 269 296 BY SIMILARITY.  
 FT DISULFID 300 312 BY SIMILARITY.  
 FT DISULFID 316 332 BY SIMILARITY.  
 FT DISULFID 335 339 BY SIMILARITY.  
 FT DISULFID 513 522 BY SIMILARITY.  
 FT DISULFID 517 530 BY SIMILARITY.  
 FT DISULFID 533 542 BY SIMILARITY.  
 FT DISULFID 546 562 BY SIMILARITY.  
 FT DISULFID 565 578 BY SIMILARITY.

```

FT DISULFID 569      BY SIMILARITY.
FT DISULFID 598      BY SIMILARITY.
FT DISULFID 625      BY SIMILARITY.
FT DISULFID 628      BY SIMILARITY.
FT DISULFID 632      BY SIMILARITY.
FT MOD RES 1141     PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250     PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68       N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573      N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631      N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661       V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 6129264583011402 CRC64;

Query Match      86.1%; Score 5865; DB 1; Length 1257;
Best Local Similarity 86.1%; Pred. No. 1.6e-304;
Matches 1082; Conservative 54; Mismatches 119; Indels 2; Gaps 2;

Qy 1 MELAALCRWGLLLALLPFGAASQVCTCTDMKRLRPASPETHLDMRLHLYOGCQVQGNL 60
Dy 1 MELAAWCRWGLLLALLPFGIAGTQVCTCTDMKRLRPASPETHLDMRLHLYOGCQVQGNL 60
Qy 61 ELTYLPNVSFLQDIOEVGYVLIJAHNQVRQVPLQRLRIVRGQLPEDNYALAVLDNG 120
Dy 61 ELTYVPANASFLQDIOEVGYVLIJAHNQVRQVPLQRLRIVRGQLPEDKYALAVLDNR 120
Qy 121 DPLNNTTPTV -GASPGGLREQLRLSLTEILKGGVLIQNRNOLCYODTILKDIHKNQL 179
Dy 121 DQONVAASTPRTPEGLREQLRLSLTEILKGGVLIQNRNOLCYODMVLKDFRKNQL 180
Qy 180 ALTLIDNRSRACHPCSPMKGSCWGESSEDCOSLRTVCAGGCARCKGLPTDCCHEQ 239
Dy 181 APVDIDNRSRACPPCAPCKDNHCWGESPEDCQILTGICTSGCARCKGLPTDCCHEQ 240
Qy 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPPEGRYTFGASCVTAC 299
Dy 241 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPPEGRYTFGASCVTTC 300
Qy 300 PYNLYSTDVGCTLVCPLNHNEVTAEDGTQCEKSCPCARVCYGLGWYIKANSKFIGI 359
Dy 301 PYNLYSTEVGCTLVCPNPNQEVTAEDGTQCEKSCPCARVCYGLGMEHURGARAITSD 360
Qy 360 TELFAGCKKIFGSLAFPEFDGDPASNTAPQVIKANSKFIGITELTYLYISAWPDSL 419
Dy 361 NVQEFDCCKIFGSLAFPEFDGDPSSGIAPLRPEQLQVFETLEEITGYLYISAWPDSL 420
Qy 420 PDLVSFQNLQVIRGILHNGAYSULTQGLGISWGLSLRLSLRSLRSLRSLRSLRSLRSLR 479
Dy 421 RDLVSFQNLRLIRGILHDGAYSULTQGLGISWGLSLRLSLRSLRSLRSLRSLRSLRSLR 480
Qy 480 VPWDLFRNPHOALLHTANRPEDE -CVGEGLACHOLCARGHCWGPPTQCVNCSQFLRGQ 538
Dy 481 VPWDLFRNPHOALLHNSRPEEDLCVSSGLVNCSLCHGHCWGPPTQCVNCSHFRLRGQ 540
Qy 539 ECVEECRVQLQPLREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPFCA 598
Dy 541 ECVEECRVKGLPREYVSDKRLPCHPECPQNSSETCFGEADQCAACAHYKDSSCVA 600
Qy 599 RCPSGVKPDLSPYMPDKFDEEGACQPCPINCNTSHSCVDLDKGCAPQORASPLTSIVAS 658
Dy 601 RCPSGVKPDLSPYMPDKFDEEGICQPCPINCNTSHSCVDLDKGCAPQORASPLTSIVAT 660
Qy 659 VGILLVVVGVWFGILIKRROQKIRKYMRLLOTELVEPLTPSGAMPNQAQRILKET 718
Dy 661 VGILLVLLVVVGVWFGILIKRROQKIRKYMRLLOTELVEPLTPSGAMPNQAQRILKET 720
Qy 719 ELRKVKVLGSAFTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGV 778
Dy 721 ELRKVKVLGSAFTVYKGIWIPDGENYKIPVAIKVLRENTSPKANKEILDEAYVMAGV 780
Qy 779 SPYVSRLLGICLSTVQLVTLQMPYGCLLDHRVNRGRGLSGDQLLNCWQIAKMSYLED 838

```

```

Dy 781 SPYVSRLLGICLSTVQLVTLQMPYGCLLDHRVNRGRGLSGDQLLNCWQIAKMSYLED 840
Qy 839 VRLVHRDLAARNVLKSPNHYKIIDFGLARLLDIDETEHADGGKVPKKNMALESILRRR 898
Dy 841 VRLVHRDLAARNVLKSPNHYKIIDFGLARLLDIDETEHADGGKVPKKNMALESILRRR 900
Qy 899 FTHOSDVMSYGVTVWELMTFCAPKPYDGIIPAREIDPLEKGRRLPOPPICITIDVYIMVKC 958
Dy 901 FTHOSDVMSYGVTVWELMTFCAPKPYDGIIPAREIDPLEKGRRLPOPPICITIDVYIMVKC 960
Qy 959 WMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPSSPMDSTFYRSLLEDDDMGDLV 1018
Dy 961 WMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPSSPMDSTFYRSLLEDDDMGDLV 1020
Qy 1019 DAESYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLPSESEAPRSLAPS 1078
Dy 1021 DAESYLVPOQGFPCDPAPGAGGMVHHRSSSTRSGGDLTLGLPSESEAPRSLAPS 1080
Qy 1079 EGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPE 1138
Dy 1081 EGAGSDVFDGDLGMAAKGLQSLPHTDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPE 1140
Qy 1139 YVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLT 1198
Dy 1141 YVNPQEVQPPPLTPEGLPPVRPAGATLERAKTSLSPKNGVVKDVFAGGAVENPEYLV 1200
Qy 1199 POGGAAQPPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLOVVPV 1255
Dy 1201 PREGTAPPHPPSPAFSPAFDNLVYWDQNSQEQGPPSPNFECTPTAENPEYLGLOVVPV 1257

RESULT 3
ERR2_MESAU
ID ERR2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERBB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RT Gene 140:251-255(1994).
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC	EMBL; D16295; BAA03801.1; --	
DR	HSSP; P11362; 1FCG.	
DR	InterPro; IPR000494; EGFR_L_domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	InterPro; IPR004019; YLP_motif.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF01030; Recep_L_domain; 2.	
DR	Pfam; PF02757; YLP; 2.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.	
KW	Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;	
KW	Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;	
KW	Proto-oncogene; Disease mutation.	
FT	SIGNAL	1
FT	CHAIN	22 1254
FT	DOMAIN	22 652
FT	DOMAIN	653 675
FT	TRANSMEM	676 1254
FT	DOMAIN	158 368
FT	DOMAIN	472 644
FT	DOMAIN	720 987
FT	NP_BIND	726 734
FT	BINDING	753 753
FT	ACT_SITE	845 845
FT	DISULFID	195 204
FT	DISULFID	199 212
FT	DISULFID	236 244
FT	DISULFID	240 252
FT	DISULFID	255 264
FT	DISULFID	268 295
FT	DISULFID	299 311
FT	DISULFID	315 331
FT	DISULFID	334 338
FT	DISULFID	511 520
FT	DISULFID	515 528
FT	DISULFID	531 540
FT	DISULFID	544 560
FT	DISULFID	563 576
FT	DISULFID	567 584
FT	DISULFID	587 596
FT	DISULFID	600 623
FT	DISULFID	626 634
FT	DISULFID	630 642
FT	MOD_RES	1139 1139
FT	MOD_RES	1247 1247
FT	CARBOHYD	68 68
FT	CARBOHYD	125 125
FT	CARBOHYD	187 187
FT	CARBOHYD	259 259
FT	CARBOHYD	530 530
FT	CARBOHYD	571 571
FT	CARBOHYD	629 629
FT	VARIANT	658 658
FT	VARIANT	659 659
FT	SEQUENCE	1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Qy	61	ELTYLPTWNASIFLQDIOEVGYVLIAHNVQRVPLQRLRIVRGQLFEDNYALAVLDNG	120
Dd	61	ELTYLPTWNASIFLQDIOEVGYVLIAHNVQRVPLQRLRIVRGQLFEDNYALAVLDNR	120
Qy	121	DPLNNTTPVTGASPGGLRELOLRSLTEILKGGVLIQRNPOLCYQDTILWKDIFHKWNOLA	180
Dd	121	DPLONVTTATGRTPEGLRELOLRSLTEILKGGVLIQRNPOLCYQDTVLWKDQVFRKNOLA	180
Qy	181	LTLDTNRSRACHPCSPWCKSGSRCWGSESSDCQSLTRTVCAAGCARCKPLPTDCCHEQC	240
Dd	181	PVDIDTNRSRACPAPACKDNHMGASPEDCQTLTGIIAPRAVPAARLARPLTDCCHQC	240
Qy	241	AAGCTGPKHSCLACLAPHNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP	300
Dd	241	AAGCTGPKHSCLACLAPHNHSIGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTTCP	300
Cy	301	YNYLSTDVGSCTLCPLPHNOBVTABDGTORCEKSKPCARVCYGLGMOYIKANSKFGIT	360
Dd	301	YNYLSTEVGSCTLCVCLPNNQEBVTABDGTORCEKSKPCARVCYGLGMEHLRGARAITSAN	360
Qy	361	ELEFAGCKKIFGSLAFIPESPDGDPASNTAPOYTKANSKFGITELTGILYISAWPDSL	420
Dd	361	IQEFAGCKKIFGSLAFIPESPDGPNSSGIAPLTPEQLOVFTELTGILYISAWPDSLH	420
Qy	421	DLSVFQNLQVIRGRILHNGAYSITLOGIGISWGLRSLRELGSGLALITHNTHLCFVHTV	480
Dd	421	DLSVFQNLRVIRGRVLHDGAYSIALQGLGIRWLGLRSLRELGSGLVLHNRNTHLCFVHTV	480
Qy	481	PWDQJFRNPHQALLHTANRPEDECVGGLGACHQICARGHCHWGPQPTQCVCNSQFLRGQEC	540
Dd	481	PWDQJFRNPHQALLHSGNPSSEEGCKLDFACYPLCAHGHCHWGPQPTQCVCNSHFLRGQEC	540
Qy	541	VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFPGPADQCACAHYKDPFCVARC	600
Dd	541	VKECRVWKGLPREVYNGKHCLPCHPECPQNSTETCTGSEADQCTACPHYKXDSFFCVARC	600
Qy	601	PSGVKPDLSYMPIWKFPDEBAGCQPCPINCTHSCVDLDDKGCPAEQRASPLTISVAVVG	660
Dd	601	PSGVKPDLSYMPIWKYPDEEGMCQPCPINCTHSCVDLDERGCPAEQRASPATSIATVVG	660
Qy	661	ILLVVVLGVVFGIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL	720
Dd	661	ILLFLVIGVVVGIILKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAMRILKETEL	720
Qy	721	RKVVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRRENTSPKANKEILDEAYVMAGVSP	780
Dd	721	RKVVLGSGAFGTVYKGIWIPDGENVKI PVAIKVLRRENTSPKANKEILDEAYVMAGLVS	780
Qy	781	YVSRLLGICLTSTVQLVTQLMYPYGLCLLDHVRENRGRIGSQDOLLNWCMIKAGMSYLEDVR	840
Dd	781	YVSRLLGICLTSTVQLVTQLMYPYGLCLLDHVRENRGRIGSQDOLLNWCMIKAGMSYLEDVR	840
Qy	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIKWMALESILRRRFT	900
Dd	841	LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEYHADGGKVPKIKWIALESILRRRFT	900
Qy	901	HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKCWM	960
Dd	901	HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKCWM	960
Qy	961	IDSECRPRFRELVEFSRMDARQRFVVIQNEDLGPASPLDSTFYRSLLEDDDDMDGLVDA	1020
Dd	961	IDSECRPRFRELVEFSRMDARQRFVVIQNEDLGPSSPLDSTFYRSLLEDDDDMDGLVDA	1020
Qy	1021	EYLVPOQGFPCDPAPGAGWVHRRHSSTSRGGDLTLGLEPSEEEAPRSPAPSEG	1080
Dd	1021	EYLVPOQGFPCDPAPGAGSTARRHRSSTSRGGELTLGMEPSCEEPSPAPSEG	1080
Qy	1081	AGSDVFDGDLGMCAKGLQSLPHDPSLQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV	1140
Dd	1081	AGSDVFEJELGMGATKGPOSISPRDLSPLQRYSEDPTLPTTETDGVVAPLACSPQPEYV	1140

OY 1141 NQPDVPRQPPSPREGPLPAARPGATLERAKTLSPGKNGVYKDVFAFGGAVENPEYLTPQ 1200  
 DB 1141 NQPEVRPQPLTPEGPLPPVRPGATLERPKTLSPGKNGVYKDVFTFGGAVENPEYLVRP 1200  
 OY 1201 GGAQPHPPSPAFDNLNYYWDQPPERCAPPSTFKGTPTAENPEYLGLDVPV 1255  
 DB 1201 GGSASOPH-PPALCPAFDNLNYYWDQPPSERGSPNTFEGTPTAENPEYLGLDVPV 1254

RESULT 4  
 ID\_EGFR\_HUMAN STANDARD; PRT; 1210 AA.  
 AC P00533; P06268; Q14225; Q9UMD8; Q9UMGS; Q92795; O00732;  
 AC O00688; Q9B2S2; Q9H2C9; Q9GX1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 DE protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Varden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Hayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RL Nature 309:418-425 (1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=95382957; PubMed=7654368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RL Mol. Reprod. Dev. 41:149-156 (1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reiter J.L., Maible N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056 (1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Garti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41 (1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RC TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schenl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Maible N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Genomics 71:1-20 (2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";  
 RL Science 224:843-848 (1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810 (1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:125-132 (1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396 (1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753 (1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:4920-4924 (1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA.";  
 RL Nature 309:270-273 (1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671 (1989).  
 RN [16]



```

QY 1141 NQDVRPQPSRPGPLPAARPGATLERAKTLSPGKNGVVKQVFAFGGAVENPEYL-TP 1199
||| : : : : : : : : : : : : : : : : : : : : : : : :
Db 1094 NQ-SVPKRPAGSVQNFVHNOPLNP-----APSRDPHYQD--PHSTAVGNPEYLNVT 1142
: : : : : : : : : : : : : : : : : : : : : : : :
QY 1200 QGGAAPOHPHPPAFSPADNLYWDQ-----DP-----PERGAPPSTFKGPTTA 1243
: : : : : : : : : : : : : : : : : : : : : : : :
Db 1143 Q-----PTCVNSTFDSPAHWAQKSGHQISLDNPDYQDDFPFKAKNGIFKGS-TA 1192
: : : : : : : : : : : : : : : : : : : : : : : :
QY 1244 ENPEYL 1249
||| ||| |||
Db 1193 ENAEYL 1198

RESULT 5
ID_EGFR_MOUSE STANDARD; PRT: 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bek/KGFR) gene.";
RL Oncogene 7:1957-1962 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Hibbs M.L.;
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=9125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413 (1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676 (1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Essinger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -1: FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -1: CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC tyrosine phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAA55587.1; -
CC EMBL; U03425; AAA17899.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSP; P11362; IFGK.
CC MGD; MGI:95294; Egfr.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00089; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU_3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
CC DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 648 670 POTENTIAL.
CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
CC REPEAT 75 300 APPROXIMATE.
CC REPEAT 390 600 APPROXIMATE.
CC DOMAIN 1028 1071 SER-RICH.
CC DOMAIN 714 981 PROTEIN KINASE.
CC NP_BIND 720 728 ATP (BY SIMILARITY).
CC BINDING 747 747 ATP (BY SIMILARITY).
CC ACT_SITE 839 839 BY SIMILARITY.
CC DISULFID 190 199 BY SIMILARITY.
CC DISULFID 194 207 BY SIMILARITY.
CC DISULFID 215 223 BY SIMILARITY.
CC DISULFID 219 231 BY SIMILARITY.
CC DISULFID 232 240 BY SIMILARITY.
CC DISULFID 236 248 BY SIMILARITY.
CC DISULFID 251 260 BY SIMILARITY.
CC DISULFID 264 291 BY SIMILARITY.
CC DISULFID 295 307 BY SIMILARITY.
CC DISULFID 311 326 BY SIMILARITY.
CC DISULFID 329 333 BY SIMILARITY.
CC DISULFID 506 515 BY SIMILARITY.
CC DISULFID 510 523 BY SIMILARITY.
CC DISULFID 526 535 BY SIMILARITY.
CC DISULFID 539 555 BY SIMILARITY.
CC DISULFID 558 571 BY SIMILARITY.
CC DISULFID 562 579 BY SIMILARITY.
CC DISULFID 582 591 BY SIMILARITY.
CC DISULFID 595 617 BY SIMILARITY.
CC DISULFID 620 628 BY SIMILARITY.
CC DISULFID 624 636 BY SIMILARITY.
CC MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```









```

Db      536 CNIYDGEFRFENGSCVCDPQCEKMEDEGLLTCHGPGDNCTKSHFKDGNPCVKECPD 595
Qy      603 GVKPDLISYMPDWKPPDEEGACQPCPINCTHSCVDLDDKGC-----PABQRASPL 651
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      596 GLQCANSF--IFKYADPDRECHPCPNCTQCGNGPTSHDCIYYPWTHGSHSTLPQHAR-TPL 652
Qy      652 TSIVSAVV-CILLVVVLGVVGLILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOA 710
Db      653 --IAAGVIGGLFVLIVGLTFAVYVRNKSIK-KKRALRRFL-ETELVEPLTPSGTAPNOA 708
Qy      711 QMRILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDE 770
Db      709 QLAILKETELRKVKVLGSGAGFTYKGIWIPDGENVKIPVAIKVLRNTPSPKANKETLDE 768
Qy      771 AYNMAGVGSYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRNKRGLSGODLLNMCWQIA 830
Db      769 ALIMASMDHPLVRLGLVCLSPITQLVTLQMPHGLLEYVHEHKDNIGSQLLNMCWQIA 828
Qy      831 KGSYLEDELVRHLDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMA 890
Db      829 KGMYLEERLVRHLDLAARNVLKSPNHVKITDGLARLLDIDETEHADGGKVPKIKMA 888
Qy      891 LESILRRRTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITD 950
Db      889 LECIHYRKFTHQSDVMSYGVTVVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITD 948
Qy      951 VYMIMVKMIDSECRPRPRELVSFSESRMARDPQRFVVIQNE-LGPASPLDSTFYRSL 1009
Db      949 VYVWVKMIDADSRPKFKELAAEFSESRMARDPQRFVVIQNE-LGPASPLDSTFYRSL 1008
Qy      1010 EDDMDGLVDVAEYLVPOQGFCDPAPAGGWHHRSSSTRSGGDDTLTGLEPSEEE 1069
Db      1009 DEEDLEMDVAEYLVPOQGFCDPAPAGGWHHRSSSTRSGGDDTLTGLEPSEEE 1056
Qy      1070 APRS-----PLAP-SEGAGSDVDFDGLGMAKGLQS 1100
Db      1057 TPMSGNQFYVYDGGFAEAGQSVYPRAPTSTIEAPVAQAQTAEIFDDSCNGTLRKPVA 1116
Qy      1101 LPHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVQNDVVRPQPPSPR 1153
Db      1117 PHVQEDSSQRYSDPTVFAPERSPRGELDEEGVMTMRDKPQOEVLYNPVE----- 1167
Qy      1154 EGPLPAARPAAGATLERAKTILSPCKGVVVDVAFAGVAVENPEYLTQGGGAPOPHPPA- 1212
Db      1168 ENPFVSR-----KNGDLQ-----ALDNPYHNASNG-----PPKAE 1199
Qy      1213 -----FSPAFDNLVYWDQDPPPERGA--PPSTP 1237
Db      1200 DEYVNEPLYNTFANTLGRAEYLNKILNSMPEKAKAFDNPDIWNHSLPRSTLQHPDYL 1259
Qy      1238 KGTP-----AENPEYL 1249
Db      1260 QEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ID ERB4_RAT
AC Q62956; Q922N7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN ERB4 OR TYRO-2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE=98221155; PubMed=9553078;
RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,

```

```

RA Marchionni M.A., Kelly R.A.;
RT "Neuregulins promote survival and growth of cardiac myocytes.
RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT ventricular myocytes.";
RL J. Biol. Chem. 273:10261-10269(1998).
RN [2]
RP SEQUENCE OF 848-901 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RL Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RN Neuron 6:691-704(1991).
RN [3]
RP SEQUENCE OF 1031-1198 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT "Expression of neuregulins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration.";
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC RECEPTORS (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC HEART.
CC -1- LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AF041838; AAD08899.1; -
EMBL; U52531; AAC53051.1; -
HSSP; P11362; 1FGK.
InterPro; IPR000494; EGFR_L domain.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR002174; Furin-like.
InterPro; IPR001245; Tyr_pkinase.
InterPro; IPR004019; YLP_motif.
Pfam; PF00757; Furin-like; 1.
Pfam; PF00069; pkinase; 1.
Pfam; PF01030; Recep_L_domain; 2.
Pfam; PF02757; YLP; 2.
PRINTS; PR00109; TYRKINASE.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00261; FU; 4.
SMART; SM00219; TyrKc; 1.
PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 25 POTENTIAL
FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.
FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 652 675 POTENTIAL.
FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).

```

```
FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT DOMAIN 718 985 PROTEIN KINASE
FT NP_BIND 724 732 ATP (BY SIMILARITY).
FT BINDING 751 751 ATP (BY SIMILARITY).
FT ACT_SITE 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA; 146957 MW; D944BB0996A08B41 CRC64;

Query Match 42.6%; Score 2900; DB 1; Length 1308;
Best Local Similarity 44.6%; Pred. No. 6.8e-147;
Matches 601; Conservative 192; Mismatches 394; Indels 160; Gaps 28;

Qy 1 MELA-ALCRWGLLL--ALLPPGAASQVCTGTDMLRLPASPEHLDMLRLHYQCQVQ 57
Db 1 MKLATGLWVWGSLLVAARTVQPSASQSVCACTENKLSLSLEQQYRALRYENCEVVM 60

Qy 58 GNLELTYPNASHLSFLQDIOEVGYVLIANOVQVPLQRLRIVRGQTQLEDNYVALV 117
Db 61 GNLEITSIEHRNDSLFIRSIEVGYVVALNQRYFLPLENRLIRIINGTKLYEDRYALAIF 120

Qy 118 DNGDPLNNTTPTVGTASPGGLRELOLRSLTEILKGGVLIQRPQCYODTILWKDIFHKN 177
Db 121 LNYRKDGNF-----GLQELGLKNLTELINGVGYVDQNKFLCYADTIHQDILVNPW 171

Qy 178 QLAULTIDTNRSRACHPCSPMKSCRGESSEDCQSILTRTVACAGG-ARCKGLPTDCC 236
Db 172 PSNMTLVSTIGSSGCGRCHKSCGTG-RCWGPTEHNCQTLTRTVCAEQCDGRCVGPVSDCC 230

Qy 237 HEQCAAGCTGPKHSDCLACLFHNSGICELCPALVTNTDTFSPMPEGRYTFGASCV 296
Db 231 HRECAAGCGSPKDTDCFACMNFNSGACVTCQPTFVYNPTTFQLEHNFNAKYTYGAFCV 290

Qy 297 TACPYNYLSTDVGSCTLLVCLPHNDEVTAEDGTQRCCKSKPCARVCVYGLGMQYIKANSKF 356
Db 291 KKCCHNFV-VDSSSCVACAPSSKMEV-EENGIKMKCPCTDICPRACDGI GTGSLMSAQTV 348
```

RESULT 8  
ID XMRK\_XIPMA  
AC P13386;

STANDARD; PRT; 1167 AA.

```
Qy 357 IGITELBFAGCKKIFGSLAFLPESFDGDPASNTAPOVIKANSKFIGITELTYLYISAWP 416
Db 349 DSSNIDKFINCTKINGNLIFLVGTIHGDPYNAIDAIDPEKLVNRTVREITGFLNIQWP 408

Qy 417 DSLPDLVSFQNLQVIRGRIHNGAYSILTLQGLGTSWLGRLSRLRELGSGLAIHNNHLCF 476
Db 409 PNMTDFSVFSLVITIGRVLYSGLSLILKQOQITSLQFQSLKEISAGNIYITONSILCY 468

Qy 477 VHTVPDOLERNPHOALLHTANRPEDECVEGLACHOLCARGHCWGCPGTPCVCNSQFLR 536
Db 469 YHTINWTTLFTSVNQRIVRNRRANCTAGMVCNHLCSNDGCMGPGPDQCLSCRRFSR 528

Qy 537 GQECVEBCRVLQGLPREYVNAHCLPCHPEQCP-QNGSVTCFGEADOCVCAHYKDPFF 595
Db 529 GKICIESCNLYDGEFREFENGSI-CVECDSCQEKMEDGLLTCHGPGDNCTCKSHFDPGN 588

Qy 596 CVARCPGVPDLSYMPIWKPDPDEGACQPCPINCTHSCVDLDDKGC-----PA 644
Db 589 CVEKCPDVLQANSF--IFKYADQDRCHPCHPNCTQCGNPTSHDCIYYPMTGHTLPO 646

Qy 645 EQRASPLTSTVSAVV-GILLVVLGVVFGILIKRRQOKIRKYVMRRLLOETELVEPLTPS 703
Db 647 HAR--TPL--TAAGVIGLGFILVIMALTFAVTVRRRSIK-KRALRLRFL-ETELVEPLTPS 701

Qy 704 GAMPNQAQMRILKETELRKVKVGLSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPPA 763
Db 702 GTPAQNQAQLRLKETELRKVKVGLSGAFGVYKGIWVPEGETVXIPVAIKILNETTGPKA 761

Qy 764 NKEILDAYVMAGVGSYPVSRLLGICLTSTVQLTQMLPYGCLLDHVRNGRSGQLL 823
Db 762 NVEFMDREALMASVDHPHLVRLGLVCLSPITQLVTQMLPHGCLLLEYVHEHKDNIQSULL 821

Qy 824 NWCQIAKAGMSYLEDVRLVHRDLAARNVLKNSPHNVKITDPLGLARLLDIDETEVHADGCK 883
Db 822 NWCQIAKAGMYLEERDLVHRDLAARNVLKNSPHNVKITDPLGLARLLDEGEKEYNADGCK 881

Qy 884 VPIKWMALLESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQ 943
Db 882 MPIKWMALCEHYRKFTHQSDVWSYGVTVWELMTFGKPYDGIIPREIPDLLEKGERLPQ 941

Qy 944 PPICTIDVYIMVWCWIMIDSECRPRFRELVSFMRMARDPORFVVIQNEED-LGPAASPLDS 1002
Db 942 PPICTIDVYIMVWCWIMIDDSRPFKELAAEFMRMARDPORIYVIOGDDRMKLPSPNDS 1001

Qy 1003 TFVESLLEDDMDGLVDAEYLVPOQGFCCDP-----APCA 1039
Db 1002 KFNQLLDEEDLMDMAEYLVLP-QAFNIPPIYTSRTRIDSNRSEIGHSPPPAYTPMS 1060

Qy 1040 GGMVHHRRSSSTRSGGGLTLGLEPSEEAAPRSP LAPSEGAGSDVFDGLMGAAKGLQ 1099
Db 1061 GSQFVYQDGGFATQGG--MMPYTAITSTIPEAPVA--QCATAEFDDSCNGTLRKPV 1115

Qy 1100 SLPHDPSPLQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVQDVRPQPPSP 1152
Db 1116 VPHQVEDSSTQRYSDPTVFAPERNRAELDEEYMTPMHDKPKQEYLNPIVE----- 1167

Qy 1153 REGPLPAARPAGATLERAKTLPSCPKNGVKDVFAGFAGAVENPEYLTPOGGAAPQHPBP 1212
Db 1168 -ENPFVSR-----KNGDLQ-----ALDNPETHSASSG-----PPKA 1198

Qy 1213 -----FSPAFDNLNYDQDQPPPERGA--PFST 1236
Db 1199 EDEVNEPLYLNTFTNALGNAEYMKNSLLSVPEKAKAFADPNPDYNNHSLPFRSTLQHPDY 1258

Qy 1237 FKGTPT-----AENPEYL 1249
Db 1259 LQESTYKFKQNGRIRPIVAENPEYL 1285
```



```
Db 587 PQCIPRCPHGLGDGDLT-LWKYADKMGQCPCHONCTQCGSPGLSGCRGD-IVSHSSL 644
Qy 654 IVSAVGLLVVGLVGLIKRQOKIRKYTHRRLLQETELVEPTTPSGAMNQAMR 713
Db 645 AVGLVGLLITVIVALLIVLLRRRIK-RKRTIRCLLQEKELVEPLTPSGAQNAPFLR 703
Qy 714 ILKETELRKVKVLGSGAGFTYKGIWIPDGENVAPIVAIKVLRNTSPKANKETLDEAYV 773
Db 704 ILKETEFKORVLGSGAGFTYKGLMNPDPGENIRIPVAIKVLRNTSPKANKETLDEAYV 763
Qy 774 MAGVGSYVSRLLGICLTSTVQLVTLMPYGLCLDHRNKRGLSGDGLNMCQIAKGM 833
Db 764 MASVDHPCVRLGICLTSAVLVTLMPYGLCLDHRNKRGLSGDGLNMCQIAKGM 823
Qy 834 SYLEDVRLVRDLAARNVLKSPHVKITDGLARLLDIDETEHADGGKVPYIKWMALES 893
Db 824 NYLEERHLVRDLAARNVLLKNPHVKITDGLSKLLTADEKEYQADGGKVPYIKWMALES 883
Qy 894 ILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYM 953
Db 884 ILQWYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKELIASVLENGERLPQPPICTIEVYM 943
Qy 954 IMVKCMWIDSECRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1013
Db 944 IILKCMWIDPSSRPRFRELSEFSEFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDD 1000
Qy 1014 MGLVDAAEYLVPOGFCFDPDPAPGAGGVWHRHSSTRSGGDLTLGLEPSEEAAPS 1073
Db 1001 --DVVDADVLLPYKRI-----NRQGS-----E 1021
Qy 1074 PLAPSEAGSDVDFDGLMGAAKGLQLPHTDPSPLQRYSDPTV-PLPSETDGVAPLT 1132
Db 1022 PCIPPTGH-----PVRENSITLNTSDPTQNALEKDLGDH----- 1056
Qy 1133 CSPOEYVNPQDVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPGKNG 1179
Db 1057 -----EYVNPQSTSSRLSDIYNPNVEDLTDGMPVSLSSQEAETNFSRPEYLTNQNS 1111
Qy 1180 VKVDVAFAGGAVENPEYVLTPOGGAAPQHPPPAPAFADNLYYWDQPPERGAPPSFTKG 1239
Db 1112 L-----PLVSSGSMDDPDY---QAG-----YQAAP-----LPQTGALTNGMF 1146
Qy 1240 TPTAENPEYLG 1250
Db 1147 LPAENLEYLG 1157

ID_ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21850;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9008234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=9328282; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL tyrosine kinase.";
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTA.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29366; AAA35790.1; -.
CC EMBL; M34309; AAA35979.1; -.
CC EMBL; S61953; AAB26935.1; -.
CC PIR; A36223; A36223.
CC HSP; P11362; 1FGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; -.
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
CC PROSITE; PS00109; PROTEIN KINASE_TYR; FALSE NEG.
CC PROSITE; PSS00011; PROTEIN KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
CC SIGNAL 1 19
CC CHAIN 20 1342
CC DOMAIN 20 643
CC TRANSMEM 644 664
CC DOMAIN 665 1342
CC DOMAIN 709 966
CC NP_BIND 715 723
CC BINDING 742 742
CC ACT_SITE 834 834
CC DISULFID 186 194
CC FT 190 202
CC BY SIMILARITY.
CC -----
CC POTENTIAL.
CC RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC PROTEIN KINASE.
CC ATP (BY SIMILARITY).
CC BY SIMILARITY.
CC BY SIMILARITY.
```

```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 556 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGGVYIEKDKLCHMDTIDWRDIDVRDARDAEIVKDKGR
SC -> GQFMVPSGLTPQPAQDQWYLLDDDDPRLLTLSASK
VPVTLAAV (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
E -> G (IN REF. 2).
E -> G (IN REF. 2).
Query Match 34.9%; Score 2375.5; DB 1; Length 1342;
Best Local Similarity 40.0%; Pred. No. 5.3e-119;
Matches 525; Conservative 197; Mismatches 457; Indels 135; Gaps 34;
SEQUENCE 1342 AA; 148097 MW; 7201E7E66CA374BD CRC64;

```

```

QY 10 GLLALLPPGAA--STQVCTGTDMKLRPASPTHLDMLRHLHYQCGVQVQGNLELTVLPT 67
DB 11 GLLFSLARGSEVGNQAVCPGLTNGLSVTCDAENQYQTLKYLCYCEVWGNLEIVLTH 70
QY 68 NALSFLQIDQVQGVYLAHQVROVPLQRURIVRGTLQFEDNYALAVLDNGDPLNNTT 127
DB 71 NADLSFLQIREVTGYLVAMNEFSTLPLNLRVVRGTQVYDGKFAIFVM----LNYNT 125
QY 128 PVTGASPGCLRELQRLSLTEILKGGVLIQPNPOLCYQDTILWKDIFHKNNQLALTIDTN 187
DB 126 ----NSSHALRQLRLTQLTEILSGGVYIEKDKLCHMDTIDWRDIDVRDRD--AEIVVKD 178
QY 188 RSRACHPCSPMKGSRWCMSSEDCOSLRTRVCAGGC-ARCKGPLPTDCCHOCACAGCTG 246
DB 179 NGRSCPPCHEVCKG-RCWFGSEDCQTLTKTICAPQCNCHGFCGPNPQCHDECACGCSG 237
QY 247 PKHSOCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVCATCPYNLST 306
DB 238 PQDTDCFACRHFNDSGACVPRCPQPLVYNKLTFLQLEPNPHTKYQYGVGVCAVCPHFV-V 296
QY 307 DVGSCTLVCPLNQVETADGTQRCCKSPCARVCYGLGMQVIKANSKF--IGTELE- 363
DB 297 DQTSVVRACFPDQMEYD-NKGLMKCEPCGGLCPKACEGTG-----SGSRFQTVDSNIDG 350
QY 364 FAGCKKIFGSLAFSPESFGDPSNATAPQVIKANSKFIGITLTGYLYISAMPDSIPDLIS 423
DB 351 FVNTCKILGNLDFLTGLNGDPWHKIPALDPKLANVFRVREITGVLTNQSPPHMNF 410
QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLGISWGLRLSLRELGLGLALIHNNHLICFVHTVPW 482
DB 411 VFSNLTITGGRSLYNGRFSLLIMKNLNTVSLGFRSLKEISAGRIVISANRQLCYHSLNW 470

```

```

QY 483 DQLFRNPHQALLHTA-NRPEDECVGEGELACHQICARGHCWGPGPTQCVNCSQFIRGQECV 541
DB 471 TKVLRGPTBERLDIKHNRRPRDCVAEGKCDPLCSCGGCNGPGPGQCLSCRNYSGGVCV 530
QY 542 EECRVLOGLPREVYNARHCLPCHPEQOPONGSVTCFGEADOCVACAHYKDPPECVACRP 601
DB 531 THCNFLNGEPREFAEAECSFCHPECMEGTATCNGSGSDTCAQCAHFRDGPCHVSSCP 590
QY 602 SGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGCAPABQRA----SPLTSTVSA 657
DB 591 HGVLG--AKGPIYKPDVQNECRPCHENTCQGGKGPBLODCLGQTLVLIGKTHLTMAITV 648
QY 658 VVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOSTELVEPLTPSGAMPNQOMRILK 716
DB 649 IAG--LVVIFMLGGTFLYWRGRRIQNKRAMRRYLERGESIEPLDPS-ERANKVLARIFK 705
QY 717 ETELRRKVKVLSGCAFGTVVKGWIGIPGENVKIPIVAIKVLENTSPKANKELDEAYVMAG 776
DB 706 ETELRLKVLGSGVFGTVHKGWIPGESIKIPVCIKVIEDKSGROSFOAVTDHMLAIGS 765
QY 777 VGSPPVSRLLIGICLTSTVOLVTQMPYGCILDHVRENRRGLSGODLLNMCWIAKMSYL 836
DB 766 LDHAHIVRLGLCPGSSQLVTOYLPGLSLLDHVRQHRGALGPQLLLNNGVQIAKMYVL 825
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIKWMALESILR 896
DB 826 EEHGMVHRNLAARNVLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPKIKWMALESIFH 885
QY 897 RRFTHOSDVSQVTVWELMTFGCAKDYDIPAREIDPDLKGBRLPOPPCTIDVYIMV 956
DB 886 GKXTHOSDVSQVTVWELMTFGAEPYAGLRLAEVDPDLLEKGBRLAQPOQCTIDVYIMV 945
QY 957 KCMWIDSECRPRELVSFSEFMRARDPQRFVVIQNEDLGPA---SPLDSTFYRSLLEDD 1013
DB 946 KCMWIDENIRPTFKELANEFTMRADPPRYLVIKRES-GGCIAPGPEPHGLTKNKLLEEVE 1004
QY 1014 MGDVDAEYLVPOQGFCDPAPGAGMVHHHRSSRSSTSGGDDTLGLSP-SEEBAPR 1072
DB 1005 LEPELDDLDLEAEE------NLATTLGSALSPLVGTILNRPRGSQ 1045
QY 1073 SPLASEGAGSVDFDGLGCAAKGLQSLPTHD-PSPLQRYSEDPTVPLP-----SETD 1125
DB 1046 SLLSPSSGY-MPMNQNLGESCQESAVSGSSERCPRVSLH-----PMRPGCLASES 1098
QY 1126 GYVA-----PLTCSPOPE---YVNPQDVVRPQPPSPREGP----- 1156
DB 1099 GHVTGSEAELOEKVSMCRSRSRSPRPRGDSAYHSQRHSLLTPTVPLSPGLEEDVNG 1158
QY 1157 --LPAARPAGATLERAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAPOP 1207
DB 1159 YVMPDTHLKGTPSSREGTLLSVGLSSVLGTEEEDE-EEVEYMNRRRRHSP-P 1209
QY 1208 HPPAFSPAFDNLVYWD-----QDPPERGAPPSTFTGTPTAENPEYL 1249
DB 1210 HPPRPSLEELGYEYMDVGSLSASLGTSQSCPLHPVIMPTAGTTTDEDEYEM 1263

```

## RESULT 10

```

ID ERB3 RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=96096535; PubMed=8522190;  
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 RT recombinant protein.";  
 RT Gene 165:279-284 (1995).  
 RN [2]  
 RN REVISIONS TO 85; 513 AND 565.  
 RP Hellyer N.J., Koland J.G.;  
 RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RN SEQUENCE OF 922-1097 FROM N.A.  
 RP STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
 RX MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neuroligins and their putative receptors, ErbB2 and  
 RT ErbB3, is induced during Wallerian degeneration.";  
 RT J. Neurosci. 17:1642-1659 (1997).  
 CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAk.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL; U29339; AAC28498.2; -;  
 DR EMBL; U52530; AAC53050.1; -;  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase\_1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 5.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 19 POTENTIAL;  
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.  
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 644 662 POTENTIAL.  
 FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 183 259 CVS-RICH.  
 FT DOMAIN 707 964 PROTEIN KINASE.  
 FT NP\_BIND 713 721 ATP (BY SIMILARITY).  
 FT BINDING 740 740 ATP (BY SIMILARITY).  
 FT ACT\_SITE 832 832 BY SIMILARITY.  
 FT DISULFID 186 194 BY SIMILARITY.  
 FT DISULFID 190 202 BY SIMILARITY.  
 FT DISULFID 210 218 BY SIMILARITY.  
 FT DISULFID 214 226 BY SIMILARITY.  
 FT DISULFID 227 235 BY SIMILARITY.

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SQ	SEQUENCE	1339	AA; 147545	MM; 0AA5F2402BBDFDIE CRC64;

Query Match 33.8%; Score 2303.5; DB 1; Length 1339;  
 Best Local Similarity 40.3%; Pred. No. 3.6e-115;  
 Matches 518; Conservative 173; Mismatches 435; Indels 159; Gaps 35;

QY	3	LAALCRWGLLLALLPPGAA---	STOYCTGTDKMLRLPASPETHDMLRLHYQSCOVVQGN	59
DB	7	LQVLC----	FLLSLARGSEMGNSQAVCPGTNLGLSVTGADNVOYQTLKYKEVVMGN	62
QY	60	LELYLPTNASLSFLQDIOEVGYVLI	AHNQVRQVPLRLRIVRGTQLFEDNYALAVLDN	119
DB	63	LEIVLTGHNADLSFLQWIREVTGYV	LVANNEFVPLPLNLRVVRGTQVYDGKFAIFM--	120
QY	120	GDPLNNTPTVTGASPGGLRELQ	LSRSTLTKGVLQIRNPOLCYQDTILWKDIFHNQNL	179
DB	121	---LNYNT---	NSSHALLRQLKFTQLTEILSGVYIEKNDKLCMDTIDWRDVRVR--	170
QY	180	ALTLIDTNRSRACHPCSKGRCW	GESSEDCQSLTRTVACGCG-ARCKGPLPTDCCHE	238
DB	171	GAEIVVKNNGANCPPEVCKG-	RCWGPDPDCQILTKTICAPQCRCRCPGPNQCCHD	229
QY	239	QCAAGCTGPKHSDCLACHFNHSG	ICELHCPALVTYNTDTTFESMPNPEGRYTFGASCVTA	298
DB	230	ECAGGCGSPQDTCFACRRFND	SGACVPRCPPLVYNKLTFFQLEPNHTKYQYGGVCVAS	289
QY	299	CPNYLSTDVGSCTLVCP	LHNOEVTADGTQCEKSKPCARVCYGL--GMQYIKANSKF	356
DB	290	CPHNFF-VDQTFECVRACPPD	KMEVD-KHGLKMCPCGGLCPKACEGTGSGSRYTVDSSN	347
QY	357	IGITELEFAGCKIFGSLAFLP	ESPFGDPSNAPQYIKANSKFIGITELTYLVTISAWP	416
DB	348	ID----GFVNCTKILGNLDF	LITGLNVDPMHKIPALDPEKLNVRFRVREITGLNLTQSWP	403
QY	417	DSLPLDSVFQNLQVIRGRILH	NGAYS-LTLOGLIGISWLGRLSRLRELGSGLALIHNTLC	475
DB	404	PHMHFSVFSNLTITIGRSL	NRGFSLLIMKNLNVTSLGFRSLKEISAGRVYISANQQLC	463
QY	476	FVHTVPWQDLFRNPQOALLH	TANR-PEDECVGEGLACHOLCARGHGWGPPTQCVNCSQF	534
DB	464	YHSLNMTLLRGPSEERLDI	KYDRPLGELAGKVCVDPLCSSSGGCGWGPQGLSCRY	523
QY	535	LRGQECVEECRVLQGLPREY	VNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDP	594
DB	524	SREGVCVTHCNFLQGEPRF	VHEAQCFSCHPECLPMEGTSTCNGSGSGSDACARCAHFRDGP	583
QY	595	FCVARCPGKVPDLSPYMPI	WKFPDEEGACQCPINCTHSC--VDLDKDGKPAQRASPLT	652

```
Db 584 HCVNSCPHGILG--AKGIPIKYPAQNECRPCHENCHENCTGCGNPELQDCLGQAEVLMSPKH 641
QY 653 SIYSAVVGILLVVLGVVFGILIKRROOKIR-KYTMRLRLQETELVPLTPSGAMPNQAO 711
Db 642 LVIAVTG--LAVILMILGSGFLYWRGRIQNKAMRYLREGSIEPLDPS-EKANKVL 698
QY 712 MRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVLRENTSPKANKEIIDEA 771
Db 699 ARIFKETELRKVLKVGSGVFGVHKGIWIPGESIKIPVCIKVIEDKSGRQSQAVTDHM 758
QY 772 YVMAGVSPVSRLLGLICLTSTVOLVTQMLPYGCLLDHVNRGRGLSQDILLNMCQIAK 831
Db 759 LAVGSLDHAHIVRLGLCPGSSSLQVLYQLPLGSLLDHVHKQHRETLQPOLLNNGVQIAK 818
QY 832 GMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLDIDETEXHADGKGKPIKKNAL 891
Db 819 GMYLLEHSMVHRDLARNVWLKSPSQVQVADFGVADLLPPDDKQLLHSEAKTPIKKNAL 878
QY 892 ESILRRRFTHQSDVMSYGVTVWELMTFGAKPYOGIPAREIPDLLEKGERLPQPPICTIDV 951
Db 879 ESIHFGKYTHQSDVMSYGVTVWELMTFGAEPYAGLRLAEIPDLLEKGERLAQPOICTIDV 938
QY 952 YMIWVKCWMIDSECRPFRELVSFSEMRADPQRFVIVQIENEDLQSPASPLDSTFYRSLLED 1011
Db 939 YMIWVKCWMIDENIRPTFKELANEFTRMARDPPRYLVIKRAS-GPGTP--PAAEPSVLTT 995
QY 1012 DDMGDLVDAEYLVPOQGFCCPDAPGAGGVVHRRSSSTRSGGDLTLGLEPSEE--- 1068
Db 996 KEL-----QEALEPEL-----DLDLDLEAESEGLA 1021
QY 1069 -----EAPRSPLAPSEG-----AGSDVDFGDLGCMGAAGLQSLPT 1103
Db 1022 TSLGSALLPTGTLTRPGSQSLSPSSGYMPNQSSIGAECLDSAVLGGREQFSRPSISL 1081
QY 1104 HDSPSLORYSEDPVPLPSETDGV-----APL-----TC-----SPQPE-----YVNPQ 1143
Db 1082 H-PIPRGR-----PASESEGHVTSSEALQKSVCRSRSRSPRGRDGSAYHSQR 1133
QY 1144 DVRPQPPSPREG-----LPAARPAGATLERAKTLP-SGKGVV-----KDVF 1185
Db 1134 HSLTPTVPLSPGLEEDGNGYWPHTLARGASSREGTLSSVGLSVLGTDEED-- 1191
QY 1186 AFGAVENPEYLTPOGGAAPQPHPP 1210
Db 1192 -----EEYEMNRKRGSP-PRPP 1209

RESULT 11
EGFR_DROME STANDARD; PRT; 1426 AA.
AC P0412; O61601; Q9W2G0; P81868;
DT 13-AUG-1987 (Rel. 05, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)
DE (Gurken receptor) (Tortoise protein) (Drosophila relative of ERBB).
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN 11
RP SSOURCE FROM N.A. (ISOFORMS TYPES I AND II).
RX MEDLINE=94350209; PubMed=8070664;
RA Clifford R., Schubach T.;
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals
RT that several genetically defined classes of alleles cluster in
RT subdomains of the receptor protein.";
RL Genetics 137:531-550(1994).
RN 12
RP REVISIONS.
```

```
RA Clifford R., Schubach T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN 13
RP SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=85124611; PubMed=2982499;
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;
RT "The Drosophila EGF receptor gene homolog: conservation of both
RT hormone binding and kinase domains.";
RL Cell 40:599-607(1985).
RN 14
RP SEQUENCE FROM N.A. CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RC STRAIN=Oregon-R; TISSUE=Embryo;
RX MEDLINE=87002474; PubMed=3093080;
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;
RT "Alternative 5' exons and tissue-specific expression of the
RT Drosophila EGF receptor homolog transcripts.";
RL Cell 46:1091-1101(1986).
RN 15
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.
RX MEDLINE=99102120; PubMed=9882502;
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;
RT "Several levels of EGF receptor signaling during photoreceptor
RT specification in wild-type, Ellipse, and null mutant Drosophila.";
RL Dev. Biol. 205:129-144(1999).
RN 16
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunko/ B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Ke Z., Kennison J.A., Ketchum K.A.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Wei M.-H., Ibegwan C.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2165-2195(2000).
RN 17
RP SSOURCE OF 959-1078 FROM N.A.
RC STRAIN=Daekwanryeong;
RX MEDLINE=85137938; PubMed=2983232;
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;
RT "A Drosophila genomic sequence with homology to human epidermal
```







```
Db 603 IADCGYISAYK--FDNRCKICHPECR-----TCNGAGADHCOECVHVVRDQGHCVCSEC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNYNDGVGRECHATCDGCTGPKDITIGACTTCNLAIIINNDATVRCRLKDDKDCPD- 713
Qy 609 SYMPIWKF--PDEGACOP-----CPI-----NCTH-----632
Db 714 QY--FWEVHFQEGSLKPLAGRAVCRKCHPLCELCCTNYGHEQVCSKCTHYKRQCET 771
Qy 633 -----SC-----VDLDKG-----641
Db 772 ECPADHYTDEORECFQRHPECNGCTGPGADDDCKSCRNFKLFANETGTYVYNSTMFNCTS 831
Qy 642 -CPAEOR-----ASPLTS-----IVSAVVGILLVVVLGVVGI 673
Db 832 KCPLMRHVNQYTAIGPYCAASPRSSKITANLDVNMFIITGAVLVPTICILCV--T 889
Qy 674 LIKROQKIRKYT--MRLLQETELVEPLTPSGAMPNOAMRILKTELKRVKVLGSGAF 731
Db 890 YICQKQAKKETVMTWALSGCESEPLRFSNTGANKLURIVKDAELRGVGLGMGAF 949
Qy 732 GTVYKGIWIPDGENVKIPVAIKVRENTSPKANKEILDEAYVMAGVSPYVSRLLGICLT 791
Db 950 GRVYKGVWVPSGENVKIPVAIKELKSTGAESSEEFLEAYIMASEEHVNLKLLAVCMS 1009
Qy 792 STVOLVTOLMPYGCLLDHVRENRRGLSGODLLNWCQIAKMSYLEDVRLVHRDLAARNV 851
Db 1010 SQMMLITQLMPLGLCLLDVVRNRRDKIGSKALLINWSTQIAKMSYLEEKRLVHRDLAARNV 1069
Qy 852 LVKSPNHVKITDFGLARLLDDIDETEHADGGKGVPIKMWALESLRRRTHQSDVWSYGV 911
Db 1070 LVQTPSLVKITDFGLAKLLSSDSNEYKAAGGKMPKWLALCEIRNRVFTSKSDVWAFGV 1129
Qy 912 VWELMTCAKPYDGIAREIIPDLLEKGBRLPOPICTIDVYVMIMVKWMIDSECRPRFRE 971
Db 1130 IWELLTFQORPHENIPAKDIPDLIEVGLKLEQPEICSLDIYCTLLSCHWLDAAEMPTFKQ 1189
Qy 972 LVSEFSRWARDPQRFVWJQNEDLG--PASPLDSTFYRSLLEDD-----DMGLVDVAEYLV 1026
Db 1190 LTTVFAEPADPGRYLAIPGDKFTRLPA-----YTSQDEKDLIRKLAPTDDGSEALAK 1242
Qy 1027 QQGFCEPDPAACAGMVHHRSSSTRSGGDLTGLPLPSREAP-----RSPLAPSEG 1080
Db 1243 PDDYLOPKAAFGPS-----HRTDCT-----DEMPKLNRYCKDPSKNKSS 1281
Qy 1081 AGSDVFDG---DLGMAAGKLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQP 1137
Db 1282 TGDDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCQPGP 1319
Qy 1138 EYVNPQVDRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGAVENPEYL 1197
Db 1320 NNNNMN-----NPNQNNMAAVGAAGY-----DLIGVPVSDVNPYL 1358
Qy 1198 ----TPQGAAPQH-----PPAFSP-AFONLYYWD 1224
Db 1359 LNAOTLVGESPIPTQTIGIPVMGGPGTMEVKVPMGPSEPTSSDHSYND 1408

RESULT 12
ERBB ALV
ID -ERBB ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DT Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB
OS Avian leukosis virus.
OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN (1)
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSSP; P11362; IFCK.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00069; kinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
DR TRANSFAR; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.7%; Score 1749.5; DB 1; Length 634;
Best Local Similarity 52.3%; Pred. No. 4.3e-86;
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

Qy 587 CAHYKDPFPFCVAPCSGVKPDLSYMPIWKFPDEGACOPCPINCTHSCVDLDDKGCPEAQ 646
Db 3 CAHFDGCHVCVAKCPAGVLGENDTL-VWKYADANAVCQLCHPCNTRCGKPGLEGCP--- 58
Qy 647 RASPLTSTVSAAV-GILLVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGA 705
Db 59 NGSKTPSIAAGVVGGLLCLVVGVLGIGLYLRRR-HIVKRTLRLRLQERLEPLTPSGE 117
Qy 706 MPNOAMRILKTELKRVKVLGSGAFVYKGIWIPDGENVKIPVAIKVRENTSPKAN 765
Db 118 APNQAHRLILKETEFKVKVILGSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKAN 177
Qy 766 EILDEAYVMAGVSPYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGODLLN 825
Db 178 EILDEAYVMASVDNPHVCRLLGICLTSTVQLITQMLPYGCLLDYIREHKDNIGSYLLN 237
Qy 826 CQIAKMSYLEDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDDIDETEHADGGKVP 885
Db 238 CVQIAKGMVLEERLVRDLAARNVVKTPQHVKITDFGLAKLLGADEKEYHAEGKVP 297
Qy 886 IKWVALESILRRRTHQSDVWSYGVTVWELMTFGAKPYDGIAREIIPDLLEKGBRLPQP 945
```

183 TVLDFASNLSSCPKCHPNCTEDHCWGAGEQNQTLTKVICAQCSGRGRGVKVPSDCCCHNQ 24  
 240 CAAGCTGPKGHSCLACLFNHSIGTICELHCPALVYNTDTTFESMNPBGRYTFGASCVTAC 29  
 243 CAAGCTGPRESDCLACRFRDATTCKTCTPLVLYNPTTYQMDVNPBGKYSFGATCVREC 30  
 300 PNYLSTDVGSCTLVCPHINQEVTAEDGTORCEKCKPCCARVCYGLGMOTIKANSKFIGI 35  
 303 PHNVYVTDHGSVRSNCNTDTYEV..ENGVRKCKKCDGLCSKVCGNGIGIGELKGILS..INA 36  
 360 TELE..FAGCKKIFGSLAFIPESPDGPDASNTAPQYIKANSKFIGITELTYLVISAWPDS 41  
 361 TNIDSPKCTKINGDVSILVPAFLGDAFTKTLPLDPKLDVFTVKESIGFLLIQAWPDN 42  
 419 LPDLSPQNQLVIRGRILHNGAYSILTQGLGSIWGLSLRSLRELGSGLALIHNNTHLCFVH 47  
 421 ADLYAFENLEIIRGTRKHGGYSUAVNVLKIQSLGLRSLKEISDGDIAIMKNKNLCYAD 48  
 479 TVPMDQLFRNPHOALLHTANRPEDECVGEGLAGHOLCARGHCWGPGPTQVCNCSOFLRQG 53  
 481 TNWRSLFATOSQTKTIQNNRKNKDCTADRHVCDPLCSVCGMGPGPFHCFSCRRFFSRQK 54  
 539 ECVBECRVQLGLPREYVVARHCLPCHPECPONG...SVTCFPGPEADOCVACAHYKDPFF 59  
 541 ECVKQCNILQGEPEFEREDSKLUPCHSECLVQNSTAYNTTCSGPGPDHCKMCAHFDGPH 60  
 596 CVARCPSGVPDLSYMPIWKFPDEEGACOPCPINCHTSCVDLDDKGCPAEQRASPLTSTV 65  
 601 CVKACPAVLGENDTL..VMKYADANAVCOLCHPNCTRGCCKPGLEGCP...NGSKTPSTA 66  
 656 SAVV..GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTP 70  
 657 AGVVGLLCVVVGIGLYLRRR..HIVKRKTRLLRQLQERLVEPLTP 70

Search completed: July 22, 2003, 08:46:09  
Job time : 20.2304 secs

[illegible]

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 37.9774 Seconds  
(without alignments) 4403.399 Million cell updates/sec

Title: SEQ4-325-339-12  
Perfect score: 6814  
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGIDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 . Gapext 0.5

Searched: 908470 seqs. 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Genesec\_101002.\*

1:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1980.DAT.*
2:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1981.DAT.*
3:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1982.DAT.*
4:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1983.DAT.*
5:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1984.DAT.*
6:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1985.DAT.*
7:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1986.DAT.*
8:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1987.DAT.*
9:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1988.DAT.*
10:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1989.DAT.*
11:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1990.DAT.*
12:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1991.DAT.*
13:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1992.DAT.*
14:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1993.DAT.*
15:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1994.DAT.*
16:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1995.DAT.*
17:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1996.DAT.*
18:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1997.DAT.*
19:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1998.DAT.*
20:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA1999.DAT.*
21:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA2000.DAT.*
22:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA2001.DAT.*
23:	/SID\$2/cgcdata/genesec/genesecp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query			ID	Description
	Score	Match	Length		
1	6734	98.8	1255	21	Human heregulin 2
2	6734	98.8	1255	22	Human tyrosine kin
3	6734	98.8	1255	22	HER2 transgene pla
4	6734	98.8	1255	22	Human HER2 (ErbB2)
5	6728	98.7	1255	17	AAW01151
6	6728	98.7	1255	20	HER-2/neu protein.
7	6728	98.7	1255	21	Human HER-2/neu on
8	6728	98.7	1255	21	Human HER-2/neu pr
9	6728	98.7	1255	22	Amino acid sequenc
10	6728	98.7	1255	22	Human HER-2/neu pr
					HER2/neu amino aci
					AAG988267

11	6728	98.7	1255	23	AAE24067	Human Her-2/neu pr
12	6728	98.7	1255	23	AAE20479	Human Her-2/neu pr
13	6728	98.7	1255	23	AAE51143	Human Her-2/neu pr
14	6728	98.7	1255	23	AAU77114	Human Her-2/neu pr
15	6685	98.1	1433	14	AAU39568	Sequence of c-erbB
16	6564	96.3	1223	23	AAU98923	Human breast cancer
17	6411	94.1	1220	21	AAU321208	Human HER-2/neu pr
18	5941.5	87.2	1256	21	AAE211199	Rat HER-2/neu prot
19	5941.5	87.2	1256	23	AAE51144	Rat Her-2/neu onco
20	5914.5	86.8	1256	21	AAE31206	Mouse Her-2/neu pr
21	5914.5	86.8	1256	22	AAE62860	Amino acid sequenc
22	5914.5	86.8	1256	23	AAE51151	Mouse Her-2/neu on
23	4814	70.6	919	21	AAE31203	Human HER-2/neu fu
24	4814	70.6	919	23	AAE51148	Her-2/neu extracel
25	4064.5	59.6	920	23	AAE51152	Mouse Her-2/neu ex
26	4064.5	59.6	920	23	AAE51153	Mouse Her-2/neu ex
27	3698	54.3	712	21	AAE31204	Human HER-2/neu fu
28	3698	54.3	712	23	AAE51149	Human HER-2/neu ex
29	3552	52.1	782	18	AAE19764	Her-2-GM-CSF immuno
30	3550	52.1	653	21	AAE31200	Extracellular HER-
31	3550	52.1	653	23	AAE51145	Human Her-2/neu on
32	3512	51.5	645	22	AAE60408	Human ErbB2 oncopr
33	3512	51.5	645	22	AAE51593	Human ErbB2 extrac
34	3447	50.6	951	21	AAE44993	DC8scfv-erbB2EC fu
35	3344	49.1	624	11	AAE08222	Extracellular port
36	3146	46.2	1210	21	AAE19259	Amino acid sequenc
37	3146	46.2	1210	21	AAE50616	Human EGF receptor
38	3146	46.2	1210	23	AAE33019	Human Her-1 protei
39	3146	46.2	1210	22	AAE50768	Human epidermal gr
40	3146	46.2	1210	22	AAE68420	Amino acid sequenc
41	3105	45.6	1210	23	ABE51768	Human epidermal gr
42	3084	45.3	583	23	AAE20493	Human protein for
43	3084	45.3	587	23	AAE20481	Human protein for
44	3083	45.2	589	23	AAE20484	Human protein for
45	3083	45.2	600	23	AAE20482	Human protein for

## ALIGNMENTS

## RESULT 1

AA92620  
ID AA92620 standard; Protein; 1255 AA.

AC AAY92620:

DT 10-AUG-2000 (first entry)

Human hercullin 2 (Her2)

XX	
KW	Herregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW	self-protein; cancer; breast cancer; prostate cancer;
KW	cell-associated peptide antigen; foreign epitope.

OS Homo sapiens.

XX	FH	Key
----	----	-----

FT	Key	Location/Qualifiers
	Domain	1..173

```

FT
21.47
/label= N-terminal

```

FT	/note= "mature polypeptide"
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

FT	Region	5..25
5..25	5..25	5..25

```
FT /label= insertion_region
ET /label= "insertion_region"
```

ET	Region	/note= "suitable for foreign epitope insertion"
ET	50	72

```

ET region
59../3
/label= insertion region

```

```

11 /label= insertion_region
FT /note= "suitable for foreign epitope insertion"

```

FT	Region	103..117	7'note= suitable for foreign epitope insertion

```
FT      /label= insertion_region
```

/note= "suitable for foreign epitope insertion"

FT	Region	149..163
FT		

```

Et Et /label=
Et Et /note= "

```

FT Region /label= Cysteine\_rich\_domain  
210..224  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Region 250..264  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Domain 324..483  
/label= Ligand\_binding\_domain  
FT Region 325..339  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Region 369..383  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Region 465..479  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Domain 484..623  
/label= Cysteine\_rich\_domain  
FT Region 579..593  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Domain 624..654  
/label= Transmembrane\_domain  
FT Region 632..652  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Region 653..667  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Domain 655..1010  
/label= Tyrosine\_kinase\_domain  
FT Region 661..675  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Region 695..709  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Region 710..730  
/label= insertion\_region  
/note= "suitable for foreign epitope insertion"  
FT Domain 1011..1235  
/label= C-terminal\_domain  
WO200020027-A2.  
13-APR-2000.  
05-OCT-1999; 99WO-DK00525.  
05-OCT-1998; 98DK-0001261.  
20-OCT-1998; 98US-0105011.  
(MEBI-) M & E BIOTECH AS.  
Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;  
Gautam A, Birk P, Karlsson G;  
WPI: 2000-349917/30.  
N-PSDB; AAA09455.  
Inducing immune responses to weakly immunogenic, tumor associated  
peptide antigens for the treatment of breast and prostate cancer  
Claim 62; Page 193-198; 220pp; English.  
This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
Her2 can be used in the claimed method as an autovaccine to induce a CTL  
response. Subdominant CTL epitopes, antibody binding regions and  
cysteine residues involved in disulfide bonds are preserved in the  
immunogenized forms. Regions suitable for the insertion of foreign T  
helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
CC cell-associated peptide antigens (PA) such as those associated with  
CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
CC The method comprises effecting simultaneous presentation by antigen  
CC producing cells (APCs) of the animals immune system of: (1) at least 1  
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
CC B-cell group derived from the cell-associated PA; and (2) at least 1  
CC first T-helper cell group which is foreign to the animal. Analogues of  
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
CC part of all known and predicted CTL and B-cell epitopes of the respective  
CC PA and including at least one foreign T helper epitope are also claimed.  
CC The method is used to treat prostate, prostate/breast or breast cancer  
CC when the PA is human PSM, FGF8b and Her2, respectively.  
XX  
SQ Sequence 1255 AA;  
Query Match 98.8%; Score 6734; DB 21; Length 1255;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1240; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASQTGCTDMKRLPASPETHLDMLRHLYQGCVVQGNL 60  
DB 1 MELAALCRWGLLLALLPPGAASQTGCTDMKRLPASPETHLDMLRHLYQGCVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIQEVQGVLIARNOVQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIQEVQGVLIARNOVQVPLQRLRIVRGTOLEFEDNYALAVLDNG 120  
QY 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHNKNOLA 180  
DB 121 DPLNNTTPTVGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILMKDIFHNKNOLA 180  
QY 181 LTLIDNRSRACHPCSPMCKGRCWGSESSDCQSLTRTVCCAGCARCKGLPTDCCHQC 240  
DB 181 LTLIDNRSRACHPCSPMCKGRCWGSESSDCQSLTRTVCCAGCARCKGLPTDCCHQC 240  
QY 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTNTDTFESMPNPGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNHSGICELHCPALVTNTDTFESMPNPGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLGMQVIKANSKFITG 360  
DB 301 YNYLSTDVGSCTLVCPHNEVTAEDGTQRCCKSPCARVCYGLGMHLEHRAVTSAN 360  
QY 361 ELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLQVFTLEETGYLYISAWPDSLP 420  
DB 361 IOEFAGCKKIFGSLAPLPESFDGDPASNTAPLOPEOLQVFTLEETGYLYISAWPDSLP 420  
QY 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISMGLRSLRELGSGLAIHHNTHLCFVHTV 480  
DB 421 DLSVFONLQVIRGRILHNGAYSILTQGLGISMGLRSLRELGSGLAIHHNTHLCFVHTV 480  
QY 481 PMDQLFRNPQALLHTANRPEDECVGEGILACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540  
DB 481 PMDQLFRNPQALLHTANRPEDECVGEGILACHQLCARGHCWGPGTQCVCNCSQFLRGQEC 540  
QY 541 VEECRVLQGLPREYVNAHCLFCHPECPQNGSVTCFGPEADOCVACAHYKDPFPCVARC 600  
DB 541 VEECRVLQGLPREYVNAHCLFCHPECPQNGSVTCFGPEADOCVACAHYKDPFPCVARC 600  
QY 601 PSGVAPDLSYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660  
DB 601 PSGVAPDLSYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPORASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720  
DB 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAOMRILKETEL 720  
QY 721 RKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGSF 780  
DB 721 RKVKVLGSGAFQTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGSF 780

Qy 781 YVSRLLGICLTSTVQLTQVLTQMPYCGLLDHHVRENRRGLSGDQLLNKWCQIAKMSYLEDVR 840  
 Db 781 YVSRLLGICLTSTVQLTQVLTQMPYCGLLDHHVRENRRGLSGDQLLNKWCQIAKMSYLEDVR 840  
 Qy 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVPKWKMALESILRRFT 900  
 Db 841 LVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEVHADGKVPKWKMALESILRRFT 900  
 Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960  
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTIDVYMIWVKWM 960  
 Qy 961 IDSECRPRFRELUSEFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1020  
 Db 961 IDSECRPRFRELUSEFMRMARDPQFVVIQNEIDLGPASPLDSTFYRSLLEDGMDLVA 1020  
 Qy 1021 EYLVPQGGFFCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
 Db 1021 EYLVPQGGFFCPDPAAGAGMVHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEG 1080  
 Qy 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
 Db 1081 AGSDVFDGDLGMGAAGLQSLPTHDPSPQRYSEDPTVPLPSETDGVVAPLTCSPQPEYV 1140  
 Qy 1141 NOPDVRPQSPREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
 Db 1141 NOPDVRPQSPREGPLPAARPAGATLERAKTLSPGKNGVVKOVFAFGGAVENPEYLTPO 1200  
 Qy 1201 GGAAPQHPPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255  
 Db 1201 GGAAPQHPPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

## RESULT 2

AAE12130  
 ID AAE12130 standard; Protein; 1255 AA.

XX AC AAE12130;

XX DT 18-DEC-2001 (first entry)

XX DE Human tyrosine kinase-type receptor, HER-2.

XX KW Therapeutic compound; major histocompatibility complex; vaccine;  
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;  
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
 KW antigen presenting cell; human; tyrosine kinase-type receptor.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
 FT Region 774..782  
 FT /note= "Antigenic epitope"

XX PN WO200168677-A2.

XX PD 20-SEP-2001.

XX PF 16-MAR-2001; 2001WO-US40328.

XX PR 16-MAR-2000; 2000US-0527487.

XX XX (GENZ ) GENZYME CORP.

XX PA Nicolette CA;

XX PI WPI; 2001-616284/71.

XX DR N-PSDB; AAD19731.

XX PT Novel synthetic therapeutic compound for inducing immune response and  
 PT for use in adoptive immunotherapy, has enhanced binding to major  
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX PS

XX Claim 4; Page 63-67; 69pp; English.

CC The invention relates to synthetic therapeutic compounds (antigenic  
 CC peptides) with enhanced binding to major histocompatibility complex  
 CC (MHC) molecules and enhanced immunoregulatory properties relative  
 CC to their natural counterparts. Compounds of the invention are useful  
 CC for inducing an immune response in a subject and for use in adoptive  
 CC immunotherapy. They are useful as components of anti-cancer vaccines  
 CC and to expand immune effector cells that are specific for cancers  
 CC characterised by expression of the breast cancer antigen, HER-2.  
 CC Polynucleotides that encode peptides of the invention are useful as  
 CC hybridisation probes and as primers for the detection of genes of gene  
 CC transcripts that are expressed in antigen presenting cells (APCs), to  
 CC confirm transduction of polynucleotides into host cells. The present  
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
 CC of the invention are designed based on the HER-2 antigenic peptide  
 CC (774-782).

XX SQ Sequence 1255 AA;

Query Match 98.8%; Score 6734; DB 22; Length 1255;  
 Best Local Similarity 98.8%; Pred. No. 0;  
 Matches 1240; Conservative 4; Mismatches 11; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPETHDMLRHLHYOGCVVQGNL 60  
 Db 1 MELAALCRWGLLALLPPGAASCTGCTDMKRLRPASPETHDMLRHLHYOGCVVQGNL 60  
 Qy 61 ELTYLPTNASLSFLQDIOEVQGVYLI AHNOVRQVPLQRLIRVGRQTQFEDNYALAVLDNG 120  
 Db 61 ELTYLPTNASLSFLQDIOEVQGVYLI AHNOVRQVPLQRLIRVGRQTQFEDNYALAVLDNG 120  
 Qy 121 DPLNNTPTVGTGSPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQOLA 180  
 Db 121 DPLNNTPTVGTGSPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKDIFHKNQOLA 180  
 Qy 181 LTLDITNRSRACHPCSPMCKGSRGWSSEDCQSLTRTVCAAGGCARCKGPLTDCCHQOC 240  
 Db 181 LTLDITNRSRACHPCSPMCKGSRGWSSEDCQSLTRTVCAAGGCARCKGPLTDCCHQOC 240  
 Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 Qy 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQDQCEKSKPCARVCYGLGHQOYIKANSKFQIT 360  
 Db 301 YNYLSTDVGSCTLVCPPLHNOEVTAEQDQCEKSKPCARVCYGLGHQOYIKANSKFQIT 360  
 Qy 361 ELEFAGCKITFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETITGYLYISAWPDSLP 420  
 Db 361 IQEFAGCKITFGSLAFLPESFDGDPASNTAPLOEQLOVFETLEETITGYLYISAWPDSLP 420  
 Qy 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRSLRELGLSLIHLHNTLHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRSLRELGLSLIHLHNTLHLCFVHTV 480  
 Qy 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSQFLRQEC 540  
 Db 481 PWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGPTQCVNCSQFLRQEC 540  
 Qy 541 VEECRVLQGLPREYVVARHCLPCHPECOPONGSVTCFGEADOCVCAHYKDPFCVVARC 600  
 Db 541 VEECRVLQGLPREYVVARHCLPCHPECOPONGSVTCFGEADOCVCAHYKDPFCVVARC 600  
 Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQSRASPLTSTVSAVVG 660  
 Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPEAQSRASPLTSTVSAVVG 660  
 Qy 661 ILLVVLGVVFGVGLIKRQOKIRKYTRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720  
 Db 661 ILLVVLGVVFGVGLIKRQOKIRKYTRRLLQETELVEPLTPSGAMPNQAMRILKETEL 720



Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKPIKMALESILRRPT 900  
Qy 901 HQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPCTIDVTMIMVKCM 960  
Db 901 HQSDVMSYGVTVWELMTGAKPYDGIPIAREIPDLLEKGERLPQPPCTIDVTMIMVKCM 960  
Qy 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEIDLGPASPLDSTFYRSILLEDDMDGLVDA 1020  
Qy 1021 EYLVPOQGFPCDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFPCDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Qy 1081 AGSDVDFDGLNGAAKGLQSLDTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVDFDGLNGAAKGLQSLDTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NOPDVRPQPPSPREGPLPAARPAAGATLRAKTLSPKNGVVKDVFAGGAVENPEYLTPQ 1200  
Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLRAKTLSPKNGVVKDVFAGGAVENPEYLTPQ 1200  
Qy 1201 GGAAPQHPPPAFSPAFNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPPPAFSPAFNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 4

AAU74545  
ID AAU74545 standard; Protein; 1255 AA.  
XX  
AC AAU74545;  
XX  
DT 23-APR-2002 (first entry)  
XX  
DE Human HER2 (ErbB2) polypeptide.  
XX  
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
KW glial disorder; astrocytal disorder; hypothalamic disorder;  
KW glandular disorder; macrophagal disorder; epithelial disorder;  
KW stomal disorder; blastocoealic disorder; inflammatory disorder;  
KW angiogenic disorder; immunological disorder.  
XX  
OS Homo sapiens.  
XX  
PN US2002001587-A1.  
XX  
PD 03-JAN-2002.  
XX  
PF 16-MAR-2001; 2001US-0811123.  
XX  
PR 16-MAR-2000; 2000US-189844P.  
PR 05-OCT-2000; 2000US-238327P.  
XX  
PA (ERIC/) ERICKSON S.  
PA (SCHW/) SCHWALL R.  
PA (SLIW/) SLIWKOWSKI M.  
XX  
PI Erickson S, Schwall R, Sliwkowski M;  
XX  
DR WPI; 2002-163686/21.  
DR N-PSDB; ABK14058.  
XX  
PT Treating tumour characterised by overexpression of epidermal growth  
PT factor receptor. ErbB or cancer in mammal, comprises administering  
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -  
XX  
PS Example 3; Fig 7; 93pp; English.  
XX  
CC The invention relates to treating a tumour in a mammal, where the tumour

is characterised by the overexpression of an epidermal growth factor  
receptor (ErbB) and does not respond or responds poorly, to treatment  
with an anti-ErbB antibody, comprising administering to the mammal an  
anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
treating cancer or tumours of the breast, ovary, stomach, endometrium,  
salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
prostate and bladder, preferably breast cancer. The breast cancer is a  
metastatic breast cancer or an aggressive form of metastatic breast  
cancer which overexpresses ErbB2. The method is also useful for treating  
neural, glial, astrocytal, hypothalamic, glandular, macrophagal,  
epithelial, stromal, blastocoealic, inflammatory, angiogenic and  
immunological disorders. This sequence represents the human HER2 (ErbB2)  
polypeptide of the invention.  
XX  
SQ Sequence 1255 AA;  
Query Match 98.8%; Score 6734; DB 23; Length 1255;  
Best Local Similarity 98.8%; Pred. No. 0;  
Matches 1240; Conservative 4; Mismatches 11; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPETHLMLRLHYQGCQVQGNL 60  
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPETHLMLRLHYQGCQVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLRLRIVRGTLFEDNVALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNQVRQVPLRLRIVRGTLFEDNVALAVLDNG 120  
Qy 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180  
Db 121 DPLNNTTPTVGASPGGLRELQRLSLEILKGVLIQRNPOLCYQDTILWKDIFHKQNOLA 180  
Qy 181 LTLDTNRSRACHPCSPMKGSRGSESDCQSLTRTVACGGCARCKGPLPTDCCHEQC 240  
Db 181 LTLDTNRSRACHPCSPMKGSRGSESDCQSLTRTVACGGCARCKGPLPTDCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAP 300  
Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTAP 300  
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
Qy 361 ELEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVLFETLEITGYLYISAWPDSL 420  
Db 361 IQEFAGCKKIFGSLAFPLPESFDGDPASNTAPLQPEQLQVLFETLEITGYLYISAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLETLQGLGISWLGRLSRLGSLGLALIHNTHLFCFHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLETLQGLGISWLGRLSRLGSLGLALIHNTHLFCFHTV 480  
Qy 481 PWDQLFRPHQALLHTANRPEDECYEGELACHQLCARGHCWCPGTQCVCNCSQFLRGQSC 540  
Db 481 PWDQLFRPHQALLHTANRPEDECYEGELACHQLCARGHCWCPGTQCVCNCSQFLRGQSC 540  
Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAPQASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPCAPQASPLTSIVSAVVG 660  
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOHRIKTEL 720  
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAOHRIKTEL 720  
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVQSP 780  
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVQSP 780  
Qy 781 YVSRLLGLICTSTVQLVTQLMPYGLDLPHVRENRRGLSGDLNWMCMQIAKMSYLEBVR 840

Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRCGLSGSODLLNMCQIAKMSYLEYDR 840  
Qy 841 LVHROLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILRRRT 900  
Db 841 LVHROLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALSILRRRT 900  
Qy 901 HOSDVMSGYVTWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMWKCMW 960  
Db 901 HOSDVMSGYVTWELMTFCAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYMWKCMW 960  
Qy 961 IDSECRPRELVSEFSRMARDPQRFVVIQNEEDLGASPLDSTFYRSLLDDMDGLVDA 1020  
Db 961 IDSECRPRELVSEFSRMARDPQRFVVIQNEEDLGASPLDSTFYRSLLDDMDGLVDA 1020  
Qy 1021 EBYLVPOQFFCPDPAAGAGVHHRSSSTRSGGDLTLGLESEEAAPSPAPSEG 1080  
Db 1021 EBYLVPOQFFCPDPAAGAGVHHRSSSTRSGGDLTLGLESEEAAPSPAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGDLGMAAGKLSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEV 1140  
Qy 1141 NQPDVRPOPSPREGPLPAARAGATLERAKTSLPGKNGVVDVPAFGAVENTPEYLTQ 1200  
Db 1141 NQPDVRPOPSPREGPLPAARAGATLERAKTSLPGKNGVVDVPAFGAVENTPEYLTQ 1200  
Qy 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVVP 1255

RESULT 5

AAW01111  
ID AAW01111 standard; Protein; 1255 AA.  
XX  
AC AAW01111;  
XX  
DT 01-JAN-1997 (first entry)  
XX  
DE HER-2/neu protein.  
XX  
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;  
KW breast cancer; ovary cancer; colon cancer; lung cancer;  
KW prostate cancer; immunisation; tumour; vaccine; vector.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT Domain 676..1255  
FT N-PSDB; /label= Intracellular domain  
FT /note= "claimed domain, useful for immunisation"  
XX  
PN W09630514-Al.  
XX  
PD 03-OCT-1996.  
XX  
PF 28-MAR-1996; 96WO-US01689.  
XX  
PR 31-MAR-1995; 95US-0414417.  
XX  
PA (UNIW ) UNIV WASHINGTON.  
XX  
PI Cheever MA, Disis ML;  
XX  
DR WPI; 1996-455361/45.  
DR N-PSDB; AAT40739.  
XX  
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or  
PT treatment of malignancies with which the HER-2/neu oncogene is  
XX associated  
XX  
PS Claim 2; Page 56-61; 71pp; English.

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is  
CC the product of the HER-2/neu oncogene (see also AAT40739). The  
CC protein is over-expressed in various cancers, including breast,  
CC ovarian, colon, lung and prostate. The intracellular domain of the  
CC protein can be used to immunise an animal against a malignancy with  
CC which the oncogene is associated. The polypeptide can be produced  
CC in transformed host cells for use in immunisation. Alternatively,  
CC animal cells are transfected in vivo or ex vivo with a viral vector  
CC that directs expression of the polypeptide.

XX Sequence 1255 AA;  
SQ  
Query Match 98.7%; Score 6728; DB 17; Length 1255;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLLALLPFGAASQVCTGTDMLRLPASPEHLMLRLHYQGCVVQGNL 60  
Db 1 MELAALCRWGLLLALLPFGAASQVCTGTDMLRLPASPEHLMLRLHYQGCVVQGNL 60  
Qy 61 ELYLPTNASLSFLQDIQEVQGYVLIHNVQVQVPLQRLIRVGTQLPEDNYALVDNG 120  
Db 61 ELYLPTNASLSFLQDIQEVQGYVLIHNVQVQVPLQRLIRVGTQLPEDNYALVDNG 120  
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCVQDTILWKDIFHKKNOLA 180  
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCVQDTILWKDIFHKKNOLA 180  
Qy 181 LFLIDTNRSRACHPCSPCKGSRGWSESDCQSLTRTVACGACRCKGPLPTDCCHEQC 240  
Db 181 LFLIDTNRSRACHPCSPCKGSRGWSESDCQSLTRTVACGACRCKGPLPTDCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRVTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360  
Db 301 YNYLSTDVGSCTLVCPHNOEVTABDGTORCEKSKPCARVCYGLGMEHLREVRAVTSAN 360  
Qy 361 ELEFAGCKKIFGSLAFPLESFDGDPASNTAPLQPELOVFEETLEETGLYISAWPDSLP 420  
Db 361 IQEFAGCKKIFGSLAFPLESFDGDPASNTAPLQPELOVFEETLEETGLYISAWPDSLP 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGSLWGLRLSRLSGLALIHNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGSLWGLRLSRLSGLALIHNTHLCFVHTV 480  
Qy 481 PWDQLFRPHQALLHTANRPEDECYGEGLACHOLCARGHCMGPGTQCVCNCSQFLRGQEC 540  
Db 481 PWDQLFRPHQALLHTANRPEDECYGEGLACHOLCARGHCMGPGTQCVCNCSQFLRGQEC 540  
Qy 541 VEECRVLOGLPREYVYNAHRLCHPCEQPONGSVTCFGEADQCACAHYKOPFPFCVARC 600  
Db 541 VEECRVLOGLPREYVYNAHRLCHPCEQPONGSVTCFGEADQCACAHYKOPFPFCVARC 600  
Qy 601 PSGVKPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660  
Db 601 PSGVKPDLISYMPIWKFPDEEGACQPCPNCTHSCVDLDDKGPAPQASPLTSIVSAVVG 660  
Qy 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVQSP 780  
Db 721 RKVKVLSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPKANKEILDEAYMAGVQSP 780  
Qy 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRCGLSGSODLLNMCQIAKMSYLEYDR 840  
Db 781 YVSRLLGICLTSTVQLVTQMPYGCCLLDHVRNRCGLSGSODLLNMCQIAKMSYLEYDR 840



QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900  
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900  
 QY 901 HQSDVSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMIWVKCM 960  
 DB 901 HQSDVSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVYMIWVKCM 960  
 QY 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020  
 DB 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEEDLGPASPLDSTFYRSLLLEDDMDGLVDA 1020  
 QY 1021 EYLVLPQGGFFCPDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSESEAPRSLAPSEG 1080  
 DB 1021 EYLVLPQGGFFCPDPAPGAGGVVHRRSSSTRSGGDLTLGLEPSESEAPRSLAPSEG 1080  
 QY 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEV 1140  
 DB 1081 AGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOPEV 1140  
 QY 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVVNDVFAFGAVENPEYLTTPQ 1200  
 DB 1141 NOPDVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVVNDVFAFGAVENPEYLTTPQ 1200  
 QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
 DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 6

AAW92406  
 ID AAW92406 standard; Protein; 1255 AA.  
 AC AAW92406;  
 XX  
 XX  
 DT 21-APR-1999 (first entry)  
 XX  
 DE Human HER-2/neu oncogene protein.  
 XX  
 KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
 KW malignancy; treatment; tumour.  
 XX  
 OS Homo sapiens.  
 FH  
 FT Key Location/Qualifiers  
 FT Region 676..1255  
 FT /note= "region which elicits immune response"  
 XX  
 XX US5869445-A.  
 XX  
 XX 09-FEB-1999.  
 XX  
 XX 01-APR-1996; 96US-0625101.  
 XX  
 XX 01-APR-1996; 96US-0625101.  
 PR 17-MAR-1993; 93US-0033644.  
 PR 12-AUG-1993; 93US-0106112.  
 PR 31-MAR-1995; 95US-0414417.  
 XX  
 XX (UNIW ) UNIV WASHINGTON.  
 XX  
 XX Cheever MA, Disis ML;  
 XX  
 XX WPI; 1999-152835/13.  
 DR N-PSDB; AAX01912.  
 XX

XX  
 PT Use of HER-2/neu polypeptides - for eliciting an immune response to  
 PT an HER-2/neu associated malignancy, particularly for treating or  
 PT preventing tumours  
 XX  
 XX Claim 3; Column 31-38; 26pp; English.  
 PS  
 XX This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune  
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
 CC B cells to produce an immune response to the HER-2/neu protein. The  
 CC method can be used for immunisation against a malignancy in which the  
 CC HER-2/neu oncogene is associated and in the treatment of an existing  
 CC tumour, or to prevent tumour occurrence or reoccurrence.  
 XX

SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 20; Length 1255;  
 Best Local Similarity 98.6%; Pred. No. 0;  
 Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;  
 QY 1 MELAALCRWGLLLALLPFGAASTQVCTGDMKRLRASPETHLDMRLHYQGCQVVGNL 60  
 DB 1 MELAALCRWGLLLALLPFGAASTQVCTGDMKRLRASPETHLDMRLHYQGCQVVGNL 60  
 QY 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNVQVPLQRLRI VRGTQLFEDNYALVDNG 120  
 DB 61 ELTYLPTNASLSFLQDIOEVQGYVLI AHNVQVPLQRLRI VRGTQLFEDNYALVDNG 120  
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRNPOLCVQDTILWKDIFHKNNQLA 180  
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSLEILKGGVLIQRNPOLCVQDTILWKDIFHKNNQLA 180  
 QY 181 LTLIDTNRSRACHPCSPCKSGRCGSESDCQSLTRTVACGACARCKGLPTDCCHEQC 240  
 DB 181 LTLIDTNRSRACHPCSPCKSGRCGSESDCQSLTRTVACGACARCKGLPTDCCHEQC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360  
 DB 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360  
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLISAWPDSL 420  
 DB 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLISAWPDSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYS LTQGLGISWLGRLSRLSGSLALIHNTLHLCFVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYS LTQGLGISWLGRLSRLSGSLALIHNTLHLCFVHTV 480  
 QY 481 PWDOLFRNPHOALLHTANRPEDECVGEGLAHCCHOLCARGHCWGPCTOCVNCQFLRGQEC 540  
 DB 481 PWDOLFRNPHOALLHTANRPEDECVGEGLAHCCHOLCARGHCWGPCTOCVNCQFLRGQEC 540  
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
 DB 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFPCVARC 600  
 QY 601 PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPABQASPLTSIYSAVVG 660  
 DB 601 PSGVKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPABQASPLTSIYSAVVG 660  
 QY 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720  
 DB 661 ILLVVVLGVVFGILIKRQOKIRKVTMRLLQETELVEPLTPSGAMPNQAQRILKETEL 720  
 QY 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEIIDEAYVMAGVGP 780  
 DB 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEIIDEAYVMAGVGP 780  
 QY 781 YVSLLLGICLTSTVQLVTOLMPYGCCLLDHVRNRRGLSGQDLNWCMIAGMSYLEYDVR 840  
 DB 781 YVSLLLGICLTSTVQLVTOLMPYGCCLLDHVRNRRGLSGQDLNWCMIAGMSYLEYDVR 840  
 QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900  
 DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKMALESILRRFT 900

QY 901 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960  
DB 901 HQSDVMSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960  
QY 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLLEDDMGDLVDA 1020  
DB 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLLEDDMGDLVDA 1020  
QY 1021 EBYLVPOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTTLGLEPSEEEAPRSLAPSEG 1080  
DB 1021 EBYLVPOQFFCPDPAPGAGGMVHHRSSSTRSGGDLTTLGLEPSEEEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAKGLQSLPTHDPSPQLQRYSEDTVPPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPO 1200  
DB 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGGAVENPEYLTPO 1200  
QY 1201 GGAAPQPPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
DB 1201 GGAAPQPPPPAFSPAFDNLYYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 7  
ID AAB21198  
XX AAB21198 standard; protein; 1255 AA.  
AC AAB21198;  
DT 12-JAN-2001 (first entry)  
DE Human HER-2/neu protein.  
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX OS Homo sapiens.  
XX PN WO200044899-A1.  
XX PD 03-AUG-2000.  
XX PF 28-JAN-2000; 2000WO-US02164.  
XX PR 29-JAN-1999; 99US-0117976.  
XX PA (CORI-) CORIXA CORP.  
XX PA (SMIK) SMITHKLINE BEECHAM.  
XX PI Cheever MA, Gheysen D;  
XX DR WPI; 2000-505976/45.  
XX DR N-PSDB; AAA89736.  
XX PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
XX PT useful for vaccinating against breast, ovarian, colon, lung and  
XX PT prostate cancers -  
XX PS Claim 52; Fig 7; 128pp; English.  
XX CC The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
XX SQ Sequence 1255 AA;  
Query Match 98.7%; Score 6728; DB 21; Length 1255;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPFGAASSTVCTGDMKRLRASPETHLDMRLHLYQGCQVVOGNL 60  
DB 1 MELAALCRWGLLLALLPFGAASSTVCTGDMKRLRASPETHLDMRLHLYQGCQVVOGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGVYLI AHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLSFLQDIOEVQGVYLI AHNQVRQVPLQRLRIVRGTQOLFEDNYALAVLDNG 120  
QY 121 DPLNNTTVPVTGASPGGLRELQRLSRLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180  
DB 121 DPLNNTTVPVTGASPGGLRELQRLSRLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180  
QY 181 LTLIDTNSRACHPCSPMKSGRCSSESSEDCOSLTRTVCAGGCARCKGPLTDCCHEOC 240  
DB 181 LTLIDTNSRACHPCSPMKSGRCSSESSEDCOSLTRTVCAGGCARCKGPLTDCCHEOC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIT 360  
DB 301 YNYLSTDVGSCTLVCPLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360  
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVLFETLEITGYLISAMPDLSLP 420  
DB 361 IOEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOQVLFETLEITGYLISAMPDLSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELSGSLALHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELSGSLALHNNTHLCFVHTV 480  
QY 481 FWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540  
DB 481 FWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVNCSOFLRGQEC 540  
QY 541 VEECRVLOGLPREYVYNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKOPPCVARC 600  
DB 541 VEECRVLOGLPREYVYNAHCLPCHPECOPOGNSVTCFGEADOCVACAHYKOPPCVARC 600  
QY 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVSAVVG 660  
DB 601 PSGVKPDLSTYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720  
DB 661 ILLVVVLGVVFGIILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780  
DB 721 RKVKVLGSGAGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYMAGVGP 780  
QY 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRNRLGSLQDLNLCMOIAGKMSYLEDDR 840  
DB 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRNRLGSLQDLNLCMOIAGKMSYLEDDR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRPT 900  
DB 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMALLESILRRRPT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960  
DB 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYMIWVKWM 960  
QY 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLLEDDMGDLVDA 1020  
DB 961 IDSECRPRFRELVSFSEMRARDPQRFVVIQNEDELGPASPLDSTFYRSLLLEDDMGDLVDA 1020

Db 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020  
Qy 1021 EYLVPOQGFCCPPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFCCPPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGMAAGLQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVDVPAFGGAVENPEYLTPO 1200  
Db 1141 NOPDVRPQPPSPREGPLPAARPAGATLERAKTLPSPKNGVVDVPAFGGAVENPEYLTPO 1200  
Qy 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLYWQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 8

AAAY84780  
ID AAY84780 standard; Protein; 1255 AA.

AC AAY84780;

XX 08-AUG-2000 (first entry)

XX Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;  
XX tumor cell proliferation; tissue degeneration; arthropathy;  
XX bone resorption; inflammatory disease; degenerative disorder;  
XX wound healing.

OS Homo sapiens.

XX WO200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99WO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX N-PSDB; AAA14812.

XX Nucleic acid encoding an erbb 2 receptor protein designated SPLICE  
XX erbb-2, inhibitors of the protein are useful for treatment of cancer -

PS Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbb-2 receptor protein. The  
XX protein has an in-frame deletion of 16 amino acids, 2 of which are  
XX conserved cysteine residues, compared to the unspliced protein. The  
XX erbb-2 polynucleotide is used to construct probes for detecting  
XX disorders of cell transformation such as cancer. Antibodies to the  
XX protein may be used to detect SPLICE erbb-2 in a sample. Agents  
XX (e.g. antisense oligonucleotides) which inhibit the expression of  
XX SPLICE erbb-2 are useful for reducing tumor cell proliferation and  
XX treating cancer. Substances which stimulate SPLICE erbb-2 are useful  
XX for treating conditions involving damaged cells including conditions  
XX in which degeneration of tissue occurs, such as arthropathy, bone  
XX resorption, inflammatory diseases, degenerative disorders of the  
XX central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 21; Length 1255;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGILLALLPPGAASSTQVCTGDMKRLRASPETHLDMRLHYQCCVQVQGNL 60  
Db 1 MELAALCRWGILLALLPPGAASSTQVCTGDMKRLRASPETHLDMRLHYQCCVQVQGNL 60  
Qy 61 ELTYLPTNASLSFLQDIQEVQYVLIHNOVROVPLQRLRIRVGTQQLPDNEDNALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIQEVQYVLIHNOVROVPLQRLRIRVGTQQLPDNEDNALAVLDNG 120  
Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180  
Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDIFHKNNQLA 180  
Qy 181 LTLIDTNRSRACHPCSPMKGSRGWSESSEBQCSLTRTVACGACARCKGLPTDCHEQC 240  
Db 181 LTLIDTNRSRACHPCSPMKGSRGWSESSEBQCSLTRTVACGACARCKGLPTDCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMQVIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMHREVRVTSAN 360  
Qy 361 ELEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFTLEBITGYLIISAWPDSL 420  
Db 361 IQEFAGCKKIFGSLAFIPESPDGPASNTAPLOPELOVFTLEBITGYLIISAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYS LTLQIGISMLGLRSRLGSLALHNNHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYS LTLQIGISMLGLRSRLGSLALHNNHLCFVHTV 480  
Qy 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCGPPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGLAHQLCARGHCGPPTQCVNCSQFLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPPEVCVAC 600  
Db 541 VEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCACAHYKDPPEVCVAC 600  
Qy 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCINCHSCVDLDDKCPAQASPLTSIISAVVG 660  
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCPINCINCHSCVDLDDKCPAQASPLTSIISAVVG 660  
Qy 661 ILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGILIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Db 721 RKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVMAGVGP 780  
Qy 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRNRRGLSGODLLNKCQIAKMSVLEDRV 840  
Db 781 YVSRLLGICLTSTVOLVTLMPYGCCLLDHVRNRRGLSGODLLNKCQIAKMSVLEDRV 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILRRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMALESILRRRFT 900  
Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVKCM 960  
Db 901 HQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDYVMIMVKCM 960  
Qy 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020  
Db 961 IDSECRPRFRELSEFSESRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDMDGLVDA 1020  
Qy 1021 EYLVPOQGFCCPPAPGAGCMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080

Db 1021 BEYLVPOQFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080  
QY 1081 AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
Db 1081 AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
QY 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVFAFGGAVENPEYLTPO 1200  
Db 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVFAFGGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 9

AA885458  
ID AAB85458 standard; Protein; 1255 AA.  
AC AAB85458;  
XX  
DT 25-SEP-2001 (first entry)  
XX Human HER-2/neu protein.  
DE Antigen-presenting cell; immunogenic; immune response; HER-2/neu;  
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.  
XX  
OS Homo sapiens.  
XX  
PN WO200153463-A2.  
XX  
PD 26-JUL-2001.  
XX  
PF 19-JAN-2001; 2001WO-US01850.  
XX  
PR 21-JAN-2000; 2000US-0177545.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Cheever MA, Hand-Zimmermann S;  
XX  
DR WPI; 2001-476112/51.  
DR N-PSDB; AAH23392.  
XX  
PT New antigen-presenting cells, useful as vaccines for eliciting or  
PT enhancing an immune response to HER-2/neu protein, particularly useful  
XX for treating or preventing cancer, e.g. breast cancer -  
XX  
PS Claim 2; Page 41-46; 49pp; English.

XX The invention provides an isolated antigen-presenting cell, which  
CC expresses at least an immunogenic portion of a polypeptide that produces  
CC an immune response to HER-2/neu protein. The antigen-presenting cells are  
CC useful as vaccines for eliciting or enhancing an immune response to  
CC HER-2/neu protein, particularly in treating or preventing malignancies in  
CC which the HER-2/neu oncogene is associated. Specifically these are  
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,  
CC colon, lung or prostate cancers. The present sequence represents  
XX the human HER-2/neu protein (also known as p185 or c-erbB2).  
XX  
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 22; Length 1255;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

QY 1 MELAAALRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRLHYQCQVQGNL 60  
Db 1 MELAAALRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRLHYQCQVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIEVQGVYLIHNVQVPLQLRIVRGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIEVQGVYLIHNVQVPLQLRIVRGTQLFEDNYALAVLDNG 120  
QY 121 DPLNNTTVPVGASPGCLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDILFKHKNOLA 180  
Db 121 DPLNNTTVPVGASPGCLRELQRLSLTEILKGVLIQRNPOLCYQDTILWKDILFKHKNOLA 180  
QY 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGGCARCKPLPTDCCHEQC 240  
Db 181 LTLIDTNSRACHPCSPMKSGRCWGESSEDCQSLTRTVACGGCARCKPLPTDCCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTAC 300  
Db 241 AAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCYTAC 300  
QY 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360  
Db 301 YNYLSTDVGSCTLVCPHLNQEVTAEDGTORCEKSKPCARVCYGLGMHLEHREVRVTSAN 360  
QY 361 ELEFAGCKKI FGS LAF LPES PDG DPASNTAP LQEP LQVFET LEE ITGYLYI SAWPDSL P 420  
Db 361 IQEFAGCKKI FGS LAF LPES PDG DPASNTAP LQEP LQVFET LEE ITGYLYI SAWPDSL P 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLGSLALIHNTLHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRSLGSLALIHNTLHLCFVHTV 480  
QY 481 PWDQLFRPHQALLHTANRPEDECYEGGLACHQLCARGHCWGPPTQCVCNCSOFLRGQEC 540  
Db 481 PWDQLFRPHQALLHTANRPEDECYEGGLACHQLCARGHCWGPPTQCVCNCSOFLRGQEC 540  
QY 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFPFCVARC 600  
Db 541 VEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFGEADQCACAHYKDPFPFCVARC 600  
QY 601 PSGVKPDLSYMPIMKFPDEEGACQPCINCHSCYVDLDDKGPAPAEQASPLTSIVSAVVG 660  
Db 601 PSGVKPDLSYMPIMKFPDEEGACQPCINCHSCYVDLDDKGPAPAEQASPLTSIVSAVVG 660  
QY 661 ILLVVVLGVVVGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVVGILIKRQOKIRKYTMRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMAGVGSP 780  
Db 721 RKVVLGSGAGFTGVYKGIWIPDGENVKIPVAIKVIRENTSPKANKEILDEAYMAGVGSP 780  
QY 781 YVSRLLGICLTSTVOLVTQLMPYGCCLDHRVNRGRGLSQDILLNMCQIAKGMSTYLEDVR 840  
Db 781 YVSRLLGICLTSTVOLVTQLMPYGCCLDHRVNRGRGLSQDILLNMCQIAKGMSTYLEDVR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKPIKWMALLESILRRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVKPIKWMALLESILRRRFT 900  
QY 901 HQSDVWSGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPCTIDVTMIVWKCM 960  
Db 901 HQSDVWSGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPCTIDVTMIVWKCM 960  
QY 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGASPPLDSTFYRSLLDDMDGLVDA 1020  
Db 961 IDSECRPRFRELVSFSESRMARDPQRFVVIQNEEDLGASPPLDSTFYRSLLDDMDGLVDA 1020  
QY 1021 EBYLVPQOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080  
Db 1021 EBYLVPQOGFFCPDPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEAAPSLAPSEG 1080  
QY 1081 AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
Db 1081 AGSDVFDGDLGGAAGLQSLTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQPEVY 1140  
QY 1141 NOPDVRPOPSPREGPLPAARPAAGATLERAKTSLPGKNGVVDVFAFGGAVENPEYLTPO 1200



```
Db 1021 EYLVPOGFCPPAPGAGGMMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGLQSLPTHPSPLOQYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140
Qy 1141 NOPDVRPOPSPREGPLPAARPAGATLERAKTLSPGKNGVVKDVFAFGAVENPEYLTQP 1200
Db 1141 NOPDVRPOPSPREGPLPAARPAGATLERPKTLSPGKNGVVKDVFAFGAVENPEYLTQP 1200
Qy 1201 GGAAPQHPHPAFSPAFDNLYWDDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255
Db 1201 GGAAPQHPHPAFSPAFDNLYWDDPPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 11
AAE24067
ID AAE24067 standard; Protein; 1255 AA.
XX
AC AAE24067;
XX
DT 23-SEP-2002 (first entry)
XX
DE Human Her-2 protein.
XX
KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;
hyperproliferative disorder; prophylaxis; inflammation; antisense;
tumour; gene therapy; phosphorothioate backbone.
XX
OS Homo sapiens.
XX
PN WO200222636-A1.
XX
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowseert LM;
XX
DR WPI; 2002-471192/50.
XX
DR N-PSDB; AAD38904.
XX
PT Novel antisense oligonucleotide which modulates the expression of Human
Epidermal Growth Factor receptor, Her2, is useful for treating tumors
inflammation or to prevent infection in humans -
XX
PS Example 13; Page 95-107; 116pp; English.
XX
CC The invention relates to antisense compounds targetted to a nucleic
acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
conditions associated with Her2 such as hyperproliferative disorders
e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
to prevent or delay infection, inflammation and tumour formation. The
invention is also used in gene therapy. The present sequence is human
Her-2 protein.
XX
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 23; Length 1255;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLHYQGCVVQGNL 60
Db 1 MELAALCRWGLLALLPPGAASQVCTGTDMLRLPASPEHLDMLRHLHYQGCVVQGNL 60
```

Db 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERPKTLSPGKNGVVKDVFAGGAVENPEYLTPO 1200  
Qy 1201 GGAAPQHPPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYLGDDVPV 1255  
Db 1201 GGAAPQHPPPAFSPAFDNLYWQDPPPERGAPSTFKGTPTAENPEYLGDDVPV 1255  
RESULT 12  
AAE20479 ID AAE20479 standard; Protein; 1255 AA.  
XX AC AAE20479;  
XX DE 01-JUL-2002 (first entry)  
XX XX Human Her-2/neu protein.  
XX KW Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
XX KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
XX OS Homo sapiens.  
XX XX  
XX Key Location/Qualifiers  
XX Region 1021..1030  
XX /note= "Naturally processed HLA-B44-restricted epitope"  
XX W0200214503-A2.  
XX 21-FEB-2002.  
XX 14-AUG-2001; 2001WO-US41733.  
XX 14-AUG-2000; 2000US-225152P.  
XX 28-SEP-2000; 2000US-236428P.  
XX 21-FEB-2001; 2001US-270520P.  
XX (CORI-) CORIXA CORP.  
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;  
XX McNeill PD, Vedvick TS;  
XX WPI; 2002-280758/32.  
XX N-PSDB; AAD32743.  
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
XX prevention and diagnosis of cancer, preferably breast cancer -  
XX Disclosure; Page 114-117; 129pp; English.  
XX The invention relates to an isolated Her-2/Neu polypeptide composition  
XX effective for eliciting an immune response. The invention is useful for  
XX eliciting an immune response in a patient, where the patient is human  
XX leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
XX The composition is useful for the therapy and diagnosis of cancer,  
XX preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
XX and other compositions for the diagnosis, prevention and treatment of  
XX human malignancies, for stimulating and/or expanding T cells specific for  
XX Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
XX patient. The invention is useful for stimulating a T cell response in a  
XX human patient, as probe or primer for nucleic acid hybridisation, to  
XX selectively form duplex molecules with complementary stretches of the  
XX entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
XX length gene from a suitable library, and to direct expression of a  
XX polypeptide in appropriate host cells. The composition is useful in  
XX prophylactic or therapeutic applications and for the treatment of cancer,  
XX preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
XX associated malignancies. The invention is useful in gene therapy. The  
XX present sequence is human Her-2/neu protein.  
XX SQ Sequence 1255 AA;  
Query Match 98.7%; Score 6728; DB 23; Length 1255;

Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;  
Qy 1 MELAALCRWGLLALLPGGAASCTQCTGDMKRLPASPTHLMRLHYGCGVVOGNL 60  
Db 1 MELAALCRWGLLALLPGGAASCTQCTGDMKRLPASPTHLMRLHYGCGVVOGNL 60  
Qy 61 ELYLPTNASLSFLQDIQEVGYVLIHAHNOVRQVPLQRLRIVRGTLQEDNYVALAVLNG 120  
Db 61 ELYLPTNASLSFLQDIQEVGYVLIHAHNOVRQVPLQRLRIVRGTLQEDNYVALAVLNG 120  
Qy 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNNOLA 180  
Db 121 DPLNNTTPTVGASPGGLREQLRLSLTEILKGGVLIORNPOLCYQDTILWKDIFHKNNOLA 180  
Qy 181 LTLIDTNRSRACHPCSPMCKGSRGWSSEDCQSLTRTVCAAGGACARCKPLTDCCHQC 240  
Db 181 LTLIDTNRSRACHPCSPMCKGSRGWSSEDCQSLTRTVCAAGGACARCKPLTDCCHQC 240  
Qy 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300  
Db 241 AAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNPEGRTYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCITLVCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360  
Db 301 YNYLSTDVGSCITLVCPLHNOEVTAEADGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360  
Qy 361 ELBPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEITCYLYISAWPDSL 420  
Db 361 IQBPAGCKKIFGSLAFLPESFDGDPASNTAPLOEQLOVETLEITCYLYISAWPDSL 420  
Qy 421 DLSVFQNLQVIRGRIHNGAYSULTLOGLIGISWLGRLSRLGSLALIHNNTHLCFVHTV 480  
Db 421 DLSVFQNLQVIRGRIHNGAYSULTLOGLIGISWLGRLSRLGSLALIHNNTHLCFVHTV 480  
Qy 481 PWDQLFRNPHQALLHTANRPEDEBCVGEGLACHQLCARGHCHGPGPTQCVNCSQFLRGQEC 540  
Db 481 PWDQLFRNPHQALLHTANRPEDEBCVGEGLACHQLCARGHCHGPGPTQCVNCSQFLRGQEC 540  
Qy 541 VECRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600  
Db 541 VECRVLQGLPREVYNARHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600  
Qy 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPAEQASPLTSIVAVVG 660  
Db 601 PSGVKPDLSPYMPIWKFPDEGACQPCPINCTHSCVDLDDKGPAPAEQASPLTSIVAVVG 660  
Qy 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVLGVVFGILIKRQOKIRKYTMRLLOQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVWAGVSP 780  
Db 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKETILDEAYVWAGVSP 780  
Qy 781 YVSRLLGICLITSTVQLVTQMLPYGCLLDHVRNKRGLSDLLNWCQIAGMSYLEVDVR 840  
Db 781 YVSRLLGICLITSTVQLVTQMLPYGCLLDHVRNKRGLSDLLNWCQIAGMSYLEVDVR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESILRRFT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPIKWMALLESILRRFT 900  
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVWCKM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTIDVYIMVWCKM 960  
Qy 961 IDSECRPRFRELVSFESRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
Db 961 IDSECRPRFRELVSFESRWARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDDMDGLVDA 1020  
Qy 1021 EBYLVPQOQFFCDDPAPGAGGMVHRRHSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Db 1021 EBYLVPQOQFFCDDPAPGAGGMVHRRHSSTSGGGDLTLGLEPSEEEAPRSLAPSEG 1080



Db 1021 EYLVPQQFFCPDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEEAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGWAAGLQSLPTTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Db 1081 AGSDVFDGDLGWAAGLQSLPTTHDPSLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
Qy 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVKDVFAFGGAVENPEYLTTPQ 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERPKTSLSPGKNGVKDVFAFGGAVENPEYLTTPQ 1200  
Qy 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

## RESULT 13

AAWS1143

ID AAWS1143 standard; Protein; 1255 AA.

XX AC AAWS1143;

XX 17-JUN-2002 (first entry)

XX Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;

XX tyrosine kinase; receptor; c-erbB2; gene therapy.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Domain 1..653 /note= "extracellular domain"

FT Domain 676..1255

FT Domain /note= "intracellular domain"

FT Domain 990..1255

FT Domain /note= "phosphorylation domain"

XX WO200212341-A2.

XX 14-FEB-2002.

XX 03-AUG-2001; 2001WO-US24283.

XX 03-AUG-2000; 2000US-0632507.

XX (CORI-) CORIXA CORP.

XX (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

XX Cheever MA, Gheysen D;

XX WPI; 2002-241743/29.

XX N-PSDB; ABA92250.

XX Her-2/neu fusion protein for treating or preventing cancer by eliciting

XX or enhancing an immune response to the protein, has Her-2/neu

XX extracellular domain fused to Her-2/neu intracellular or

XX phosphorylation domain

XX Claim 68; Fig 7; 141pp; English.

XX

XX The present sequence is that of human Her-2/neu (p185 glycoprotein

XX or c-erbB2), an oncogenic self-protein and target for anti-cancer

XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed

XX in a variety of cancers, including breast, ovarian, colon, lung and

XX prostate cancer. Her-2/neu is a member of the tyrosine kinase

XX family of receptor-like glycoproteins. It comprises an extracellular

XX domain with homology to the epidermal growth factor receptor

XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal

XX intracellular domain that also shows homology to EGFR. Its

XX overexpression correlates with a poor prognosis in breast and

XX ovarian cancers. The invention provides Her-2/neu fusion

XX proteins, nucleic acids encoding them, viral vectors, and vaccines

XX

XX

XX

XX

XX

XX

XX

CC comprising the fusion proteins or nucleic acid molecules. In  
CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
CC protein is fused to a Her-2/neu intracellular domain or  
CC phosphorylation domain (or its DeltaPD fragment). An immune  
CC response to Her-2/neu protein is elicited or enhanced by  
CC administering the fusion protein in the form of a vaccine, or by  
CC transfecting cells of an animal ex vivo with a nucleic acid  
CC encoding the fusion protein, and delivering the transfected cells  
CC to the animal. The fusion proteins, nucleic acids, and isolated  
CC specific T-cells are useful for inhibiting the development of a  
CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
CC in a patient. T cells that specifically react with a Her-2/neu  
CC fusion protein can be used to remove tumour cells from a sample in  
CC order to inhibit the development of cancer in a patient.

XX Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 23; Length 1255;

Best Local Similarity 98.6%; Pred. No. 0;

Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;

Qy 1 MELAALCRWGLLLALLPPGAASSTOCTGTDKMLRLPASPETHLDMRLHYQCCVVOGNL 60

Db 1 MELAALCRWGLLLALLPPGAASSTOCTGTDKMLRLPASPETHLDMRLHYQCCVVOGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSFLQDIOEVQYVLIAHNQVRQVPLQRLRIVRGTLQFEDNYALAVLDNG 120

Qy 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKQIFHKQNQLA 180

Db 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWKQIFHKQNQLA 180

Qy 181 LTLIDTNRSRACHPCSPMKCGSRGSESDCQSLTRTVAGGCARCKPLPTDCCHQEC 240

Db 181 LTLIDTNRSRACHPCSPMKCGSRGSESDCQSLTRTVAGGCARCKPLPTDCCHQEC 240

Qy 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Db 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300

Qy 301 YNYLSTDVGSCTLVCPHNVQVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360

Db 301 YNYLSTDVGSCTLVCPHNVQVTAEDGTQRCCKSKPCARVCYGLGMEHLREAVTSAN 360

Qy 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEITGYLISANPDSL 420

Db 361 IQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFETLEITGYLISANPDSL 420

Qy 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRLSRLSGLALIHNTLHLCFVHTV 480

Db 421 DLSVFQNLQVIRGRILHNGAYSILTQGLGISWGLRLSRLSGLALIHNTLHLCFVHTV 480

Qy 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCGPGTQCVCNSQFIRGQEC 540

Db 481 PWDQLFRNPHQALLHTANRPEDECVEGLACHQLCARGHCGPGTQCVCNSQFIRGQEC 540

Qy 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHYKDPKPCVARC 600

Db 541 VEECRVLOGLPREYVNAHCLPCHPECPQNGSVTCFPEADQCACAHYKDPKPCVARC 600

Qy 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTTSIVSAVVG 660

Db 601 PSGVKPDLSPYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTTSIVSAVVG 660

Qy 661 ILLVVVLGVVPGIILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQWRIILKETEL 720

Db 661 ILLVVVLGVVPGIILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQWRIILKETEL 720

Qy 721 RKVKVLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVSP 780

Db 721 RKVKVLGSGAFGTGYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDEAYVWAGVSP 780



QY 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGLSGQDLNNWCMQIAKMSYLEVDV 840  
DB 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGLSGQDLNNWCMQIAKMSYLEVDV 840  
QY 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRFT 900  
DB 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRFT 900  
QY 901 HQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKWM 960  
DB 901 HQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICTIDVYIMVVKWM 960  
QY 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020  
DB 961 IDSECRPRFELVSEFSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLDDDDMDGLVDA 1020  
QY 1021 EYLVLVPOQFFCPDPAPGAGMWHRRSSSTRSGGGDLTLGLEPSEEEAPRSLAPSEG 1080  
DB 1021 EYLVLVPOQFFCPDPAPGAGMWHRRSSSTRSGGGDLTLGLEPSEEEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGDLGMAAGKQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
DB 1081 AGSDVFDGDLGMAAGKQSLPTHDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPEYV 1140  
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGAVENPEYLTPO 1200  
DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVFAFGAVENPEYLTPO 1200  
QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255  
DB 1201 GGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPSTFKGTPTAENPEYLGLDVVP 1255

## RESULT 14

AAU77114  
ID AAU77114 standard; Protein; 1255 AA.  
XX  
AC AAU77114;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Human Her-2/neu polypeptide.  
XX  
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;  
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;  
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;  
KW Hodgkin's lymphoma; T cell therapy.  
XX  
OS Homo sapiens.  
XX  
PN W0200213847-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 13-AUG-2001; 2001WO-US25408.  
XX  
PR 14-AUG-2000; 2000US-0638280.  
PR 28-SEP-2000; 2000US-0675904.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Gaiger A, Cheever MA, Hand-zimmermann S;  
XX  
DR WPI; 2002-280741/32.  
DR N-PSDB; ABK10730.  
XX  
PT Inhibiting haematological malignancy development by administering  
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide  
PT encoding the polypeptide, or antigen presenting cells expressing the  
PT polypeptide.  
XX  
PS Disclosure; Page 71-74; 74pp; English.  
XX

CC The invention relates to a method for inhibiting development of  
CC haematological malignancy in a patient by administering a polypeptide  
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
CC encoding the polypeptide. Antigen presenting cells that express the  
CC protein can also be administered. The sequences are used for inhibiting  
CC development of haematological malignancy such as acute myelogenous  
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
XX  
SQ Sequence 1255 AA;

Query Match 98.7%; Score 6728; DB 23; Length 1255;  
Best Local Similarity 98.6%; Pred. No. 0;  
Matches 1238; Conservative 5; Mismatches 12; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLPASPETHLDMRLHYQGCQVVGNL 60  
DB 1 MELAALCRWGLLLALLPFGAASQTCTGDMKRLPASPETHLDMRLHYQGCQVVGNL 60  
QY 61 ELTYLPTNASLFLQDIOEVQYVLI AHNOVROVPLRLRIVRGTOLFEDNYALAVLDNG 120  
DB 61 ELTYLPTNASLFLQDIOEVQYVLI AHNOVROVPLRLRIVRGTOLFEDNYALAVLDNG 120  
QY 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180  
DB 121 DPLNNTTPTVTCASPGGLRELQRLSLTEILKGVLIQRNPQLCYQDTILWKDIFHKNNQLA 180  
QY 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTCAGGCARCKGLPTDCHEQC 240  
DB 181 LTLIDTNRSRACHPCSPMKGRCWGESSEDCQSLTRTCAGGCARCKGLPTDCHEQC 240  
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
QY 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSKPCARVCYGLGMVYKANSKFIGIT 360  
DB 301 YNYLSTDVGSCTLVCPHNEQVTAEDGTQRCCKSKPCARVCYGLGMVYKANSKFIGIT 360  
QY 361 ELEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPELOVFTLEBITGYLISAWPDSLP 420  
DB 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTAPLQPELOVFTLEBITGYLISAWPDSLP 420  
QY 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLIGLSWGLRSRLSGSLALHNNTHLCFVHTV 480  
DB 421 DLSVFQNLQVIRGRILHNGAYS LTLQGLIGLSWGLRSRLSGSLALHNNTHLCFVHTV 480  
QY 481 PWDQFRNPQALLHTANRPEDECVGEGLCACHQCHGCHGPGPTQCVNCSOFURGOBEC 540  
DB 481 PWDQFRNPQALLHTANRPEDECVGEGLCACHQCHGCHGPGPTQCVNCSOFURGOBEC 540  
QY 541 VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600  
DB 541 VEECRVLOGLPREYVYVNAHCLPCHPECPQNGSVTCFGEADQCACAHYKDPFPCVARC 600  
QY 601 PSGVKPDLISYMPIWKPFDEBACQCPINCTHSCVDLDDKGPAPORASPLTISIVAVVG 660  
DB 601 PSGVKPDLISYMPIWKPFDEBACQCPINCTHSCVDLDDKGPAPORASPLTISIVAVVG 660  
QY 661 ILLVVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
DB 661 ILLVVVLGVVFGILIKRRQOKIRKYTWRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
DB 721 RKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDEAYVMAGVSP 780  
QY 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGLSGQDLNNWCMQIAKMSYLEVDV 840  
DB 781 YVSRLLGICLTSTVLTQMPYGCCLLDHVRNRRGLSGQDLNNWCMQIAKMSYLEVDV 840  
QY 841 LVHRDLAARNVLVKS PNHVKITDFGLARLLDIDETEHADGCKVPIKWMALLESILRRFT 900

Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRPT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVTMIMVKWM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVTMIMVKWM 960  
QY 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
QY 1021 EYLVPOQGFPCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFPCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080  
QY 1081 AGSDVFDGLGMAAGLQSLPTHDPSPLORYSEDPVPLSETDGYVAPLTCSPQPEV 1140  
Db 1081 AGSDVFDGLGMAAGLQSLPTHDPSPLORYSEDPVPLSETDGYVAPLTCSPQPEV 1140  
QY 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAFGAVENPEYLTQ 1200  
Db 1141 NQPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVVKDVPFAFGAVENPEYLTQ 1200  
QY 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255  
Db 1201 GGAAPQHPHPPAFSPAFNLVYWDQDPPERGAPPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 15  
AA39568  
ID AAR39568 standard; Protein; 1433 AA.  
XX  
AC AAR39568;  
XX  
DT 07-FEB-1994 (first entry)  
XX  
DE Sequence of c-erbB-2 tumour antigen.  
XX  
KW Tumour antigen; c-erbB-2; glycoprotein.  
XX  
OS Homo sapiens.  
XX  
PN W09316185-A.  
XX  
PD 19-AUG-1993.  
XX  
PF 05-FEB-1993; 93WO-US01055.  
XX  
PR 06-FEB-1992; 92US-0831967.  
XX  
PA (CETU ) CETUS ONCOLOGY CORP.  
XX  
PA (CREA-) CREATIVE BIOMOLECULES INC.  
XX  
PI Houston LL, Huston JS, Oppermann H, Ring DB;  
XX  
DR WPI; 1993-272889/34.  
XX  
DR N-PSDB; AAQ46083.

PT New single chain Fv polypeptide binding to C-erbB-2 tumour  
PT antigen - for imaging of treating breast or ovarian cancer etc.  
XX  
PS Disclosure; pages 48-54; 87pp; English.  
XX  
CC c-erbB-2 refers to a protein antigen expressed on the surface of  
CC tumour cells. such as breast and ovarian tumour cells, which is an  
CC approx. 200,000 mol wt. acidic glycoprotein having an isoelectric  
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents  
CC the location of a stop codon in AAQ46083.  
XX  
SQ Sequence 1433 AA;

Query Match 98.1%; Score 6685; DB 14; Length 1433;  
Best Local Similarity 98.1%; Pred. No. 0;

Matches 1231; Conservative 7; Mismatches 17; Indels 0; Gaps 0;  
QY 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPETHLDMRLHYQCGVVOGNL 60  
Db 1 MELAALCRWGLLLALLPPGAASTQVCTGTDMLRLPASPETHLDMRLHYQCGVVOGNL 60  
QY 61 ELTYLPTNASLFLQDIOEVQYVLIHNRQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120  
Db 61 ELTYLPTNASLFLQDIOEVQYVLIHNRQVRQVPLQRLRIVRGTLQFEDNYALAVLNG 120  
QY 121 DPLNNTTPTVTCASPGGLRELOLRLSLEILKGVLIQRPOLCYQDTILWKDIFHKQNQLA 180  
Db 121 DPLNNTTPTVTCASPGGLRELOLRLSLEILKGVLIQRPOLCYQDTILWKDIFHKQNQLA 180  
QY 181 LTLDTNRSRACHPCSPMKSCWGESSEDCQSILTRTVACGGACRCKGPLDCCHEOC 240  
Db 181 LTLDTNRSRACHPCSPMKSCWGESSEDCQSILTRTVACGGACRCKGPLDCCHEOC 240  
QY 241 AAGCTGPKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300  
Db 241 AAGCTGPKHSIDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300  
QY 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSFPGIT 360  
Db 301 YNYLSTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVRAVTSAN 360  
QY 361 ELEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFELEITGLYVLSAMPDSL 420  
Db 361 IQEPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFELEITGLYVLSAMPDSL 420  
QY 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLGSLGIALIHNHNLFCFVHTV 480  
Db 421 DLSVFQNLQVIRGRILHNGAYSILTLQGLIGISWGLRLSRLGSLGIALIHNHNLFCFVHTV 480  
QY 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCHGPGTQCVNCSQFIRGQSC 540  
Db 481 PWDQLFRNPHOALLHTANRPEDECVGEGGLACHQLCARGHCHGPGTQCVNCSQFIRGQSC 540  
QY 541 VEECRVLQGLPREVYNARHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600  
Db 541 VEECRVLQGLPREVYNASHCLPCHPECPQNGSVTCFGEADOCVACAHYKDPFPCVARC 600  
QY 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTISIVSAVVG 660  
Db 601 PSGVKPDLSYMPIWKFPDEEGACQPCINCTHSCVDLDDKGCAPQASPLTISIVSAVVG 660  
QY 661 ILLVVVLGVVFGIILKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db 661 ILLVVVLGVVFGIILKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQMRILKETEL 720  
QY 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGS 780  
Db 721 RKVVLGSGAFGTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVMAGVGS 780  
QY 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEVDR 840  
Db 781 YVSRLLGICLTSTVQLVTQLMPYGCLLDHVRENRLGSLQDILLNMCQIAKMSYLEVDR 840  
QY 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRPT 900  
Db 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKWMALLESILRRPT 900  
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVTMIMVKWM 960  
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVTMIMVKWM 960  
QY 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
Db 961 IDSECRPRFRELVSFEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLDDDDMGDLVDA 1020  
QY 1021 EYLVPOQGFPCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080  
Db 1021 EYLVPOQGFPCPDPAAGAGMVHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080

Qy	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Db	1081	AGSDVFDGDLGMGAAGLQSLPTHDPSPLORYSEDPTVPLPSETDGYVAPLTCSPQPEYV	1140
Qy	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVVKDVFAFGGAVENPEYLTPO	1200
Db	1141	NQPDVRFQPPSPREGPLPAARPAAGATLERPKTSLSGKNGVVKDVFAFGGAVENPEYLTPO	1200
Qy	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVAV	1255
Db	1201	GGAAPQHPHPPAFSPAFDNLYYWDQDPPPERGAPSTFKGTPTAENPEYLGLDVAV	1255

Search completed: July 22, 2003, 08:41:12  
Job time : 43.9774 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.3575 Seconds  
(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-325-339-12  
Perfect score: 6814  
Sequence: 1 MELAAALCRWGLLLALLPPGA.....TFKGTPTAENPEYLGLDVVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21:\*  
1: sp\_archea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6217	91.2	1259	6 O18735	O18735 canis famil
2	3144	46.1	1209	11 Q9QX70	Q9QX70 rattus norv
3	3115	45.7	1210	11 Q9EP98	Q9EP98 mus musculu
4	2727	40.0	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2694	39.5	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2297	33.7	1328	13 P79754	P79754 fugu rubrip
7	2029	29.8	1433	5 Q9BIH9	Q9BIH9 anopheles g
8	1871	27.5	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.5	367	11 Q8R2X1	Q8R2X1 mus musculu
10	1720	25.2	729	15 Q86712	Q86712 avian rous-
11	1718	25.2	567	15 Q86714	Q86714 avian rous-
12	1697	24.9	412	4 Q8WYV0	Q8WYV0 homo sapien
13	1653	24.3	962	15 Q64895	Q64895 avian eryth
14	1645	24.1	545	15 Q85468	Q85468 avian eryth
15	1506	22.1	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1490	21.9	643	11 Q9ERV6	Q9ERV6 mus musculu

17 1276 18.7 1193 5 Q9Y1X8  
18 1190.5 17.5 1368 5 Q23821  
19 1162 17.1 1717 5 Q26566  
20 1126 16.5 527 13 Q90836  
21 1001.5 14.7 478 11 Q9ESE0  
22 942.5 13.8 599 13 Q9PSH2  
23 906 13.3 165 4 Q14256  
24 887 13.0 176 11 Q923V5  
25 806.5 11.8 346 13 P11776  
26 778 11.4 435 5 Q8SZM1  
27 754.5 11.1 311 13 Q9P162  
28 734.5 10.8 1362 13 Q9PV24  
29 734 10.8 331 4 Q9BUD7  
30 730 10.7 1671 5 Q9NJV5  
31 723 10.6 149 6 Q9BG66  
32 692 10.2 1418 13 Q93457  
33 686.5 10.1 1368 13 Q8UM85  
34 671.5 9.9 1369 13 Q8UM86  
35 662.5 9.7 1472 5 Q9U5A8  
36 658 9.7 1412 13 Q8UM84  
37 657 9.6 1358 13 Q73798  
38 641.5 9.4 1418 13 Q8UM83  
39 632 9.3 1245 13 Q9YGH8  
40 622 9.1 1371 11 Q9QW4  
41 620.5 9.1 2144 5 Q9VD94  
42 598 8.8 987 11 Q91YM0  
43 595 8.7 935 4 Q96L35  
44 595 8.7 987 11 Q99MR2  
45 587.5 8.6 1036 4 Q07912

## ALIGNMENTS

## RESULT 1

O18735 O18735 PRELIMINARY; PRT; 1259 AA.  
AC O18735;  
DT 01-JAN-1998 (TREMBlrel. 05, Created)  
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE ErBB-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "cDNA cloning of erbB-2 from canine mammary gland";  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1; -  
DR HSP; P11362; IFGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP\_2\_domain; 1.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;

Query Match		91.2%; Score 6217; DB 6; Length 1259;
Best Local Similarity		90.9%; Pred. No. 0;
Matches 1145; Conservative 43; Mismatches 66; Indels 6; Gaps 2;		
QY	1	MELAAACRWGLLLALLPAGAASTQVCTGDMKRLRASPETHLDMRLHYQCQVVOGNL 60
DB	1	MELAAACRWGLLLALLPAGAASTQVCTGDMKRLRASPETHLDMRLHYQCQVVOGNL 60
QY	61	ELTYLPTNASLSFLQDIEVQYVLI AHNOVROVPLQRLRI VRGTQLPEDNVALAVLDNG 120
DB	61	ELTYLPTNASLSFLQDIEVQYVLI AHNOVROVPLQRLRI VRGTQLPEDNVALAVLDNG 120
QY	121	DPLNNTTPVTGASPGGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHRKNQLA 180
DB	121	DPLEGGIPAPGAAQGLRELQRLSLEILKGGVLIQRNPOLCYQDTILWKDIFHRKNQLA 180
QY	181	LTLIDNRSRACHPCSPCKGSRGSESDQSLTRIVCAGGCARCKGPLPTDCCHQC 240
DB	181	LTLIDNRSRACHPCSPCKGSRGSESDQSLTRIVCAGGCARCKGPLPTDCCHQC 240
QY	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300
DB	241	AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTSCP 300
QY	301	YNYLSTDVSGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMQVIKANSKFIGIT 360
DB	301	YNYLSTDVSGSCTLVCLPHNQEVTAEDGTORCEKSKPCARVCYGLGMHLEHREVRAVTSAN 360
QY	361	ELEFAGACKIFGSLAFLESFPGDPASNTAPLOPEQLQVFTLEBITGYLISAWPDSLP 420
DB	361	IOEFAGACKIFGSLAFLESFPGDPASNTAPLOPEQLRVFEALBEITGYLISAWPDSLP 420
QY	421	DLVSFQNLQVIRGRILHNGAYSLTQGLGISWGLRSRLRELGLGSLALHNTLHCFVHTV 480
DB	421	NLSVFQNLQVIRGRILHNGAYSLTQGLGISWGLRSRLRELGLGSLALHNRNLHCFVHTV 480
QY	481	PMDQLFRNPHOALLHTANRPEDECYEGEGVLACHOLCARGHGWGPGTQCVCNCSQFLRGQEC 540
DB	481	PMDQLFRNPHOALLHSANRPEECYEGEGVLACYP-CAHGHGWGPGTQCVCNCSQFLRGQEC 539
QY	541	VEECRVLOGLPREYVNAHCLCHPECPQNGSVTCFGEADQCVACAHYKDPDFCVARC 600
DB	540	VEECRVLOGLPREYVNDKYCLPCHSECQPNQSVTCFGEADQCVACAHYKDPDFCVARC 599
QY	601	PSGVRPDLISYPMIWKFPDDEGACQPCINCHTSCVDLDDKGCAPAEORASPLTSIVSAVVG 660
DB	600	PSGVRPDLISYPMIWKFADEEGTCQPCINCHTSCADLDEKGCAPAEORASPTVTSIAVVG 659
QY	661	ILLVVVLGVVFGILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQRILKETEL 720
DB	660	ILLAVVVLGVILIKRROQKIRKYTMRRLLQETELVEPLTPSGAMPNQAQRILKETEL 719
QY	721	RKVVLGSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 780
DB	720	RKVVLGSGAFGTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAYVMAGVGP 779
QY	781	VYSRLGLCTSTVOLVTLQMPYGLLDHVRNRCRLGSDLLNMCQIAKMSYLEDR 840
DB	780	VYSRLGLCTSTVOLVTLQMPYGLLDHVRNRCRLGSDLLNMCQIAKMSYLEDR 839
QY	841	LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESILRRFT 900
DB	840	LVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMMALESIPPRFT 899
QY	901	HQSDVMSYGVTVWELMTGAKBYDGI PAREIPDLLEKGERLPDPPICITIDVYMWKCM 960
DB	900	HQSDVMSYGVTVWELMTGAKBYDGI PAREIPDLLEKGERLPDPPICITIDVYMWKCM 959
QY	961	IDSECRPRFRELVSFSESRMARDPQRFVITQNEDELGPASPLDSTFYSRLLEDMDGLVDA 1020
DB	960	IDSECRPRFRELVAESRMARDPQRFVITQNEDELGPASPLDSTFYSRLLEDMDGLVDA 1019
QY	1021	EYLVVPOQGFPCDPAPGAGMGMVHRRSSSTRSGGDLTLGLEPSEEPKSLAPSEG 1080
DB	1020	EYLVVPOQGFPCPEPTCAGGTAAHRRHSSSTRNGGDLTLGLEPSEEPKSLAPSEG 1079
QY	1081	AGSDVFDGDLGMAKGLQSLPHTDPPSLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1140
DB	1080	AGSDVFDGDLGMAKGLQSLPHTDPPSLQRYSEDTVPLPSETDGYVAPLTCSPQPEYV 1139
QY	1141	NQPDVRPQPPSPREGPLPAARPAGATLER-----AKTILSPGKNGVVKDVFATGGAVENPE 1195
DB	1140	NQPEVWFPPLALEGLPPSRPAGATLERPKTILSPGKNGVVKDVFATGGAVENPE 1199
QY	1196	YLTPOGGAAPHPHPPAFSPAFDNLVYWDQDPPERGAAPPSTPKGTPTAENPEYGLDVPV 1255
DB	1200	YLAPGRAAPHPHPPAFSPAFDNLVYWDQDPSERGSPPSTFEGTPTAENPEYGLDVPV 1259
RESULT 2		
Q90X70		PRELIMINARY; PRT: 1209 AA.
ID	Q90X70;	
AC	Q90X70;	
DT	01-MAY-2000 (TRENBLrel. 13, Created)	
DT	01-MAY-2000 (TRENBLrel. 13, Last sequence update)	
DT	01-JUN-2002 (TRENBLrel. 21, Last annotation update)	
DE	Epidermal growth factor receptor.	
GN	EGFR.	
OS	Rattus norvegicus (Rat).	
OC	Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OC	NCBI_TaxID=10116;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RX	MEDLINE=90259888; PubMed=2342466;	
RA	Petch L.A., Harris J., Raymond W.W., Blaeband A.J., Lee D.C.,	
RA	Earp H.S.;	
RT	"A truncated, secreted form of the epidermal growth factor receptor is	
RT	encoded by an alternatively spliced transcript in normal rat tissue.";	
RL	Mol. Cell. Biol. 10:2973-2982(1990).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Petch L.A.;	
RL	Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=FISHER; TISSUE=LIVER;	
RA	Guttridge K., Dawson T.L., Earp H.S.;	
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; M37394; AAF14008.1; .	
DR	HSP; P11362; 1FGK.	
DR	InterPro; IPR000494; EGFR_L domain.	
DR	InterPro; IPR000719; Euk_pkinase.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR001245; Tyr_pkinase.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L domain; 2.	
DR	PRINTS; PR00109; TYRKINASE.	
DR	ProDom; PD000001; Euk_pkinase; 1.	
DR	SMART; SM00261; FU; 3.	
DR	SMART; SM00219; TyrKc; 1.	
DR	PROSITE; PS00107; PROTEIN KINASE ATP; 1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.	
KW	ATP-binding; Receptor; Transferrase; Tyrosine-protein kinase.	
SQ	SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B7773 CRC64;	
Query Match		46.1%; Score 3144; DB 11; Length 1209;
Best Local Similarity		50.3%; Pred. No. 3.1e-228;
Matches 643; Conservative 165; Mismatches 354; Indels 116; Gaps 28;		
QY	3	LAALCRWGLLLALLPAGA-ASTQVCTGDMKRLRASPETHLDMRLHYQCQVVOGNLE 61















```

Query Match      25.5%; Score 1739; DB 11; Length 367;
Best Local Similarity 88.0%; Pred. No. 6.2e-123;
Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

Qy 889 MALESILRRRTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICT 948
Db 1 MALESILRRRTHOSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICT 60

Qy 949 IDVTMIMVKWMDISECRPRELSEFSESRMARDPQRFVWIONEDLGPASPLDSTFYRSL 1008
Db 61 IDVTMIMVKWMDISECRPRELSEFSESRMARDPQRFVWIONEDLGPASPLDSTFYRSL 120

Qy 1009 LEDDMGDLVDAEYLVPOQGFCDPDPAPGAGMWHHRSSSTRSGGDLTLGLERPEE 1068
Db 121 LEDDMGDLVDAEYLVPOQGFCDPDPAPGAGMWHHRSSSTRSGGDLTLGLERPEE 180

Qy 1069 EAPRSLAPSGAGSDVFDGLGMAAGLQSLPHDPSLQRYSEDTVPILPSETDGYV 1128
Db 181 EAPRSLAPSGAGSDVFDGLGMAAGLQSLPHDPSLQRYSEDTVPILPSETDGYV 240

Qy 1129 APLTCSPOEYVNPQVVRPSPREGPLPAARAGATLERAKTILSPCKNGVVKDVFAFG 1188
Db 241 AFLACSPQEVYNQVEVRQPSLTPEGPPPIRPAAGATLERAKTILSPCKNGVVKDVFAFG 300

Qy 1189 GAVENPEYLTQCGAAPPHPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEY 1248
Db 301 GAVENPEYLTQCGAAPPHPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPTAENPEY 360

Qy 1249 LGLDVPV 1255
Db 361 LGLDVPV 367

RESULT 10
Q86712 PRELIMINARY; PRT; 729 AA.
AC Q86712;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Polyprotein.
GN POLYPROTEIN.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60725.1; -.
DR HSP; P03322; IAEs.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR004028; Retro M.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF02813; Retro M; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EPE1D63 CRC64;

Query Match      25.2%; Score 1720; DB 15; Length 729;
Best Local Similarity 54.8%; Pred. No. 4.7e-121;
Matches 358; Conservative 118; Mismatches 118; Indels 102; Gaps 15;

```

```

Qy 569 PQNGSVTCFGPEADQCACAHYKOPFCVARCPCGVKPDLSYMPWKFPDEEGACQPCPI 628
Db 141 PEETATPKTP--DHCMKAHFIDGPHCVKACPAAGVLENDTL-VWKADANAVCOLCHP 197

Qy 629 NCTHSCVDLDDKGPAPORASPLTSIVSAV--GILLVVLVGVVFGILIKRROOKIRKTYM 687
Db 198 NCTRGCKPGLEGCP--NGSKTPSIAAGVVGGLLCLVVGIGLGLYLR--HIVRKRTL 253

Qy 688 RLLQETELVPLTPSGAMPNOAMRILKETELKVKVVLGSGAFCTVYKGIWIPGENVK 747
Db 254 RLLQERELVPLTPSGAPNOAHRIUKETEFKVKVVLGSGAFCTVYKGLWIPGEKVK 313

Qy 748 IPVAIKVLRNTPSKANKEILDEAYVMAGVSPYVSRLLGLTCTSTVOLVTLMPYGCLL 807
Db 314 IPVAIKELREATSPKANKEILDEAYVMASVDNPRVCRLLGLTCTSTVOLITQLMPYGCLL 373

Qy 808 DHVTENRGRGSLQDLNWKMCQIAKMSYLEDLVRLVHROLAARNVLVKSNNHVKITDFGLA 867
Db 374 DYIREHKDNIGSQYLLNMCVQIAKGMNYLEERLLVHRLAARNVLVKTPOHVKITDFGLA 433

Qy 868 RLLDIDETEHADGKVPKIMWALESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI 927
Db 434 KLLGADEKEYHAEKVPKIMWALESILHRIYTHOSDVMSYGVTVWELMTFGSKPYDGI 493

Qy 928 AREIPDLLEKGERLPQPICTIDVTMIMVKWMDISECRPRELSEFSESRMARDPQREV 987
Db 494 ASEISSVLEKGERLPQPICTIDVTMIMVKWMDISDRPFERELIAEFSKWARDPPRYL 553

Qy 988 VIQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEYLVPOQGFCDPDPAPGAGMWHHR 1046
Db 554 VIQDERMHLPSPTDSKPYRTLMEEEDMEDIVDAEYLVPHQGF----- 598

Qy 1047 HRSSTRSGGDLTLGLEPSEEA PRSP-----APSEGAGSDVFDGLGMAAGLQSL 1101
Db 599 -NSPST-----SRTPLLSSLSATSNNSATNCID-----RNGOCH 631

Qy 1102 PTHDPSPLQRYSEDTVPILPSET--DGTVAPLTCPOEYVNPQVVRPQPPSPREGPLPA 1159
Db 632 PVREDSFQRYSSDPTGNFLSEIDDGFL-----PAPEYVNO--LMPKKS----- 675

Qy 1160 ARPAGATLERAKTILSPCKNGVVKDV-----AFGAVENPEYL 1197
Db 676 -----TAMVQNQIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL 715

RESULT 11
Q86714 PRELIMINARY; PRT; 567 AA.
AC Q86714;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11950;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.;
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities.";
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSP; P11362; IFGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.

```

```
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Tyrosine-protein kinase.
FT NON TER 1
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match      25.2%; Score 1718; DB 15; Length 567;
Best Local Similarity 55.4%; Pred. No. 4.6e-121;
Matches 357; Conservative 73; Mismatches 112; Indels 102; Gaps 15;

Qy 578 GPEADQCACAHYKDPFCVACRSGVKPDLISYPIWKFPDEGACQCPINCTHSCVDL 637
Db 1 GP--DHCWKCAHFIDGPHCVKACPAVLGENDTL-VKMYADANAVCOLCHPNCIRGCKGP 57

Qy 638 DDKCPABQASPLTSISAVV-GILLVVVGVGVGILIKRQKIRKYTRRRLLQETEL 696
Db 58 GLEGCP---NGSKTPSIAAGVVGGLLCLVVVGLGLGLYLR--HIVKRTLRLLQEREL 113

Qy 697 VEPLTPSGAMPNQAMRIKTELKVKVGLSGAGFTVYKGIWIPDGENVKIPVAIKVL 756
Db 114 VEPLTPSGEAPNQHRLIKTEFKVKVGLSGAGFTVYKGLWPEGEKVKIPVAIKEL 173

Qy 757 ENTSPKANKEILDEAYVMAGVSPVSRLLGICLTSTVOLVTQLMYPGCLLDHVRNKR 816
Db 174 EATSPKANKEILDEAYVMASVDNPRVCLLGLICLTSTVOLITQLMYPGCLLDYIREKDN 233

Qy 817 LGSODLLNWCQIAKGSYLEEDVRLVHRDLAARNVLKSPHHVKITDGLARLLDIDTE 876
Db 234 IGSQYLLNWCQIAKGMNYLEERLVRDLAARNVLKTPQHVKITDGLAKLLGADEKE 293

Qy 877 YHAGGKVPKIMWALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIDLE 936
Db 294 YHAGGKVPKIMWALESILRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLE 353

Qy 937 KGERLPQPICTIDVYIMVWKMWIDSECRPRFELVSEFSRMARDPQRFVVIQ-NEDLG 995
Db 354 KGERLPQPICTIDVYIMVWKMWIDADSRPKRELIAEFSKWARDPPRYLVIOGDERMH 413

Qy 996 PASPLDSTFYSLDEDDMDGLVDAEYLVPOQGFCCPDPAAGAGGVMVHHRSSRSG 1055
Db 414 LPSPTDSKFTYRLTMEEDMEDIVDAEYLVPHQGF-----NSPST--- 454

Qy 1056 GGDLTGLGPESEEEAPRSL-----APSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLO 1110
Db 455 -----SRTPLSSLSATSNATNCID-----RNCQGHVPVEDSFVQ 491

Qy 1111 RYSEDPTVPLPSET--DGYVAPLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAATLE 1168
Db 492 RYSDPTGNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 526

Qy 1169 RAKTLSPKNGVVKDVP-----AFGGAVENPEYL 1197
Db 527 ----TAMVQNOIYNNISLTAKLPMDSRYQNSHSTAVDNPEYL 566

RESULT 12
Q8MYVO
ID Q8MYVO PRELIMINARY; PRT; 412 AA.
AC Q8MYVO;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 44.7 kDa protein.
GN P31659.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN 1.
RP SEQUENCE FROM N.A.
RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,
```

```
Wan D.P., Gu J.R.;
"Novel human cDNA clones with function of inhibiting cancer cell growth.";
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
EMBL; AF318349; AAL55856.1; -.
InterPro; IPR002048; EF-hand.
InterPro; IPR000719; Euk_pkinase.
InterPro; IPR001245; Tyr_pkinase.
InterPro; IPR004019; YLP_motif.
Pfam; PF00069; pkinase; 1.
Pfam; PF02757; YLP; 2.
ProDom; PD000001; Euk_pkinase; 1.
SMART; SM00219; TyrKc; 1.
PROSITE; PS00118; EF HAND; UNKNOWN 1.
PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Hypothetical protein.
SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D2BC CRC64;

Query Match      24.9%; Score 1697.5; DB 4; Length 412;
Best Local Similarity 80.5%; Pred. No. 1e-119;
Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

Qy 889 MALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIDPLEKGERLPQPICT 948
Db 1 MALESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIDPLEKGERLPQPICT 60

Qy 949 IDVTYIMVWKMWIDSECRPRFELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSL 1008
Db 61 IDVTYIMVWKMWIDSECRPRFELVSEFSRMARDPQRFVVIQNEDELGPASPLDSTFYRSL 120

Qy 1009 LEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGGVMVHHRSSRSGGDLTLGLEPSEE 1068
Db 121 LEDDDMDGLVDAEYLVPOQGFCCPDPAAGAGGVMVHHRSSRSGGDLTLGLEPSEE 180

Qy 1069 EAPRSLAPSGAGSDVFDGLGMAAKGLQSLPTHDPSPQRYSEDTVPLPSETDGVV 1128
Db 181 EAPRSLAPSGAGSDVFDGLGMAAKGLQSLPTHDPSPQRYSEDTVPLPSETDGVV 240

Qy 1129 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPCKNGVVKDVFAG 1188
Db 241 APLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTLSPCKNGVVKDVFAG 300

Qy 1189 GAVENPEYLTPOGAAAPQ-----HPPPA---FSPAFLNL 1220
Db 301 GAVENPEYLTPOGAAALSP-----GAPPSTFKGTPTAEN 1245

Qy 1221 YYWD-QDPPER-----GAPPSTFKGTPTAEN 1245
Db 361 VVMTQCEPEQVRRSPDVSSGREGLTSAQIKRWEGPPTTSRGTCCHARN 410

RESULT 13
Q64895
ID Q64895 PRELIMINARY; PRT; 962 AA.
AC Q64895;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag, v-erb-A, v-erb-B protein.
DE Gag, v-erb-A, v-erb-B.
GN GAG, v-erb-A, v-erb-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=11861;
RN 1.
RP SEQUENCE FROM N.A.
RX MEDLINE=9020603; PubMed=1969616;
RA Bruskine A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
RT "Six amino acids from the retroviral gene gag greatly enhance the transforming potential of the oncogene v-erb-B.";
RL Oncogene 5:15-24 (1990).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; X52209; CAA36459.1; -.
```

DR	EMBL::X52211; CAA36459.1; JOINED.
DR	HSP8; P10828; 2NLL.
DR	InterPro; IPR000719; Euk_pkinase.
DR	InterPro; IPR000536; Hormone_rec_lig.
DR	InterPro; IPR001723; Stdhrmn_receptor.
DR	InterPro; IPR001245; Tyr_kinase.
DR	InterPro; IPR001628; Znf_Casteroid.
DR	Pfam; PF00104; hormone_rec; 1.
DR	Pfam; PF00069; pkinase; 1.
DR	Pfam; PF00105; zf-C4; 1.
DR	PRINTS; PR00398; STRDHORMONER.
DR	PRINTS; PR00047; STDIDFINGER.
DR	ProDom; PD000001; Euk_pkinase; 1.
DR	ProDom; PD000035; znf_Casteroid; 1.
DR	SMART; SM00430; HOI1; 1.
DR	SMART; SM00219; TyrKc; 1.
DR	SMART; SM00399; ZnF C4; 1.
DR	PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW	ATP-binding; DNA-binding; Nuclear protein; Receptor;
KW	Transcription regulation; Transferase; Tyrosine-protein kinase;
KW	Zinc-finger.
SQ	SEQUENCE 962 AA; 108320 MW; 3C5AED791E4E95CE CRC64;
Query Match 24.3%; Score 1653.5; DB 15; Length 962;	
Best Local Similarity 51.7%; Pred. No. 7.4e-116;	
Matches 358; Conservative 73; Mismatches 142; Indels 119; Gaps 18;	
Qy	541 VEECRVLQGLPRE-VYNAR-HCLP-----CHPEQC 568
Dd	: :
Dd	354 IEKQESYLAFEHYINYRKHNIHFWSKLMLKVADLRMIGAYHASRFLHMKEVCPTEL 413
Qy	569 PONGSVTCFGEADOCVACAHYKDPPPCVCAPCSGVKPDLSYMPIWFPDEEGACQCPPI 628
Dd	: :
Dd	414 PQE-----VGSP--DHCMKCAHFIIDGPHCVKCAPAGVIGENDTL-WMYADANAVCOLCHP 465
Qy	629 NCTHSCVDLDDKGCPAEQRASPLTSIVSAVV-GILLVVLGWFGILIKRRQOKIRKYTM 687
Dd	: :
Dd	466 NCTRGCCKPGLEGCP---NGSKTPTSIAGVVGGLLCLVVGLGIGLYLRRR-HIVRKRTL 521
Qy	688 RRLLOETLEVEPLTPSGAMPNQAMRIKETELRKVKLVGSGAGFTYVYGIMPIDGSNVK 747
Dd	: :
Dd	522 RRLIQERELVEPLTPSGEAPNQAHURLIKETEFKKVKVLGFAGPGTYVKGWIPEGEKVT 581
Qy	748 IPVAIKVIRENTSPKANKEILDEAYVMAGVCSPVSRLLIGLTSTVOLYTQLMPYCGLL 807
Dd	: :
Dd	582 IPVAIKELREATSPKANKEILDEAYVMASVDNPVHCVRLGIGLTSTVOLITQLMPYCGLL 641
Qy	808 DHVRENRLGQSOLLNCMQIAKGMSYLEDVRLVHRDLAARNVLKVSNNHKITDFGLA 867
Dd	: :
Dd	642 DYIREHKDNIGSQVLLNWCQIAGKNYLERHMHVHRDLAARNVLVKTPOHVKITDFGLA 701
Qy	868 RLIDIDETEHADGGKVPIKMALESILRRRPTHQSDWSYGVTVWELMTFGAKPYDGIP 927
Dd	: :
Dd	702 KQLGADEKEYHAEGGKVPIKMALESILRHITYTHQSDWSYGVTVWELMTFGSKPYDGIP 761
Qy	928 AREIPDLLKEGERLPQPPICTIDYVMTWKCMWIDSECRPFRELVEFSGRMARDPORFV 987
Dd	: :
Dd	762 ASEISSVLEGERLPQPPICTIDYVMTWKCMWGSGADSRPKFRELIATFKSMARDPPRYL 821
Qy	988 VIO-NEDLGASPILDSTFYRSLLDEDDMDGDLDVAEEYLVPOQQCFCCDPAPGAGMVVHR 1046
Dd	: :
Dd	822 VIQDERMHLPSPTDSKFYRTLMEEEDMEDIVDAEYLVPHQGFF----- 866
Qy	1047 HRSSSTRSGGDDLTLGLEPSEEAPRSLAPSEGAGSDVFDDGLGMAAKGLQSLPTHDP 1106
Dd	: :
Dd	867 -NSPST-----SRTPLLSSLSATSN-----NSATKCIDBNGH-- 898
Qy	1107 SPLQRYSEDFTVPLPSETDGYAPLTCSPQEVYNQPDVNRQPPSPREGPLPAARPAGAT 1166
Dd	: :
Dd	899 -----PVREDGFL-----PAPEYVNO--LMPKKPSTAMVNQOINYYNISLT 936

```
QY 1056 GGLTLGLPSEEEAPRPL-----APSEGAGSDVFDGLGMRAGKGLQSLPHTDPSPLQ 1110
Db 455 -----SRIPLSSLSATSNNGATCIRNGG-----H----- 481
QY 1111 RYSEDPTVPLPSETDGYVAPLTCSPOPEYVNPQDVRPQPPSPREGPLPAARPAGAT-LER 1169
Db 482 -----PVREDGFL-----PAPEYVNG--LMPKKPSTAMVQIQIYVISTATSK 523
QY 1170 AKTSLSPGNGVVKDVFAGGAVENPEYL 1197
Db 524 LPMDSRYNQ-----SHSTAVDNPEYL 544

RESULT 15
Q9WVF5 PRELIMINARY; PRT; 655 AA.
AC Q9WVF5;
DT 01-NOV-1999 (TREMELrel. 12, Created)
DT 01-NOV-1999 (TREMELrel. 12, Last sequence update)
DT 01-JUN-2002 (TREMELrel. 21, Last annotation update)
DE Epidermal growth factor receptor (Epidermal growth factor receptor
DE isoform 3).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,
RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA Maible N.J.;
RT "Alternative Transcripts from the Human and Mouse EGFR Genes Encode
RT Carboxy-Terminal Truncated Receptors.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEV/TAC;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AF124513; AAD44149.1; -.
DR EMBL; AF275366; AAG28047.1; -.
```

```
DR EMBL; AF275364; AAG28047.1; JOINED.
DR EMBL; AF275365; AAG28047.1; JOINED.
DR EMBL; AK004944; BAB23688.1; -.
DR EMBL; AK004883; BAB23641.1; -.
DR EMBL; AK004911; BAB23662.1; -.
DR MGI; MGI:95294; Egfr.
DR InterPro; IPR000494; EGFR_L domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L domain; 2.
DR SMART; SM00261; FU; 3.
KW Receptor.
SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.1%; Score 1506.5; DB 11; Length 655;
Best Local Similarity 44.5%; Pred. No. 5.2e-105;
Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LLLALLPPGAA--STQVCTGTDMLRUPASPTHLDMLRLHYQGVQVQGNLELTYLPTN 68
Db 14 LTLTALCAAGALEKKVCGQTSNRLTQLGTTFEDHFLSLQRMYNCEVVLGNLEITYQVRN 73
QY 69 ASLSFLQDIQEVQGVYLIHNRQVRQVPLQRLRIRVGTQTFEDNYALAVLDNGDPLNTP 128
Db 74 YLSFLTKTIQEVAGVLIANTVERIPLENLIQIRGNALYENTYALAILSN----- 124
QY 129 VTGASPGRLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDI----FHKNNALTLI 184
Db 125 -YGTNRCTGLRPLMRNLQEILIGAVRFSNNPILCNMDTIQRWDIVQNVFMSNMSDL--- 180
QY 185 DTRSRACHPCSPMKSGSRCHGESSEDCOSLTRTVCGAGCA-RCKGPIPTCCCHCEOCAG 243
Db 181 -QSHPSKCPKCDPSCPNCSGCGGEEENCQKTKIICAOCCSHRCRGRSPSCCHNQCAAG 239
QY 244 CTGPKHSCLACLHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTACPVNY 303
Db 240 CTGPRESCLVCQKQFQDEATCKDTCPLMLNPTTYQMDVNPGEKYSGATCKVCKPRNY 299
QY 304 LSTVGSCTLCPLHNOEVTAEADGTQCEKCKSPCARVCYGLGMQYIKANSKFIGITELE 363
Db 300 VVTDHGSVRAAGCPDYEV-EEDGIRKCKKCDGPKRCVNGIGIGEFK-DTLLSINATNIK 357
QY 364 -FAGCKIFGSLAFLPESFDGPASNTAPLOEQLOVPETLEEITGYLYISAWPDSLPL 422
Db 358 HFKYCTAISGLDHLPLVAFKGSFTRTPLDPLRELEILKTVKEITGFLLIQAWPNWTDL 417
QY 423 SVFQNLQVIRGIRLHNGAYSILTQGLGISWLGSLRSLRGLALIHHTHLCFVHTVPW 482
Db 418 HAFENLEIIRTKHQGFSLAVVGLNITSGLSLKLEISDGDVLSGNRLCYANTINW 477
QY 483 DQLFRNPHQALLHTANRPEDECVGEGGLACHOLCARGHCWGPGTQCVNCSOFLRGOECVE 542
Db 478 KKLFGTNPQKTKIMNRAEKDKAVNHVNCPLCSSEGCGWGPEDRCDVSCQNVSRGECVE 537
QY 543 ECRVLOGLPREYVVARHCLPCHPCQPNQSVTCFGEADOCVACAHVKDPPFCVACRPS 602
Db 538 KCNLEGEPRFVENSECQHPCLPQAMNITTCGRGPDNICOAHVIDGPHCVKTCPA 597
QY 603 GVKPDLSPYMTWPKFPDEBEGACQPCFINCTHSCVDLDDKGC 642
Db 598 GIMGENNTL-VWKYADANNVCHLCHANCYTCAGPGLOGC 636

Search completed: July 22, 2003, 09:01:01
Job time : 54.3575 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2304 Seconds  
(without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-325-339-12  
Perfect score: 6814  
Sequence: 1 MELALCRWGLLLALLPPCA.....TFKGTPTAENPEYLGIDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6728	98.7	1255	1 ERB2 HUMAN	P04626 homo sapien
2	5937	87.1	1257	1 ERB2 RAT	P06494 rattus norv
3	5921.5	86.9	1254	1 ERB2 MESAU	P00533 mesocricetu
4	3144	46.1	1210	1 EGFR HUMAN	P00533 homo sapien
5	3118	45.8	1210	1 EGFR_MOUSE	Q01279 mus musculu
6	2966.5	43.5	1308	1 ERB4 HUMAN	Q15303 homo sapien
7	2947	43.2	1308	1 ERB4 RAT	P02956 rattus norv
8	2685.5	39.4	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2423.5	35.6	1342	1 ERB3 HUMAN	P21860 homo sapien
10	2351.5	34.5	1339	1 ERB3 RAT	P06299 rattus norv
11	1952	28.6	1426	1 EGFR_DROME	P04412 drosophila
12	1749.5	25.7	634	1 ERBB ALV	P00534 avian leuko
13	1703	25.0	604	1 ERBB AVIER	P00535 avian eryth
14	1630	23.9	540	1 ERBB AVIEU	P11273 avian eryth
15	1595	23.4	703	1 EGFR_CHICK	P13387 gallus gall
16	1290	18.9	1323	1 LT23_CAEEL	P24348 caenorhabdi
17	1142.5	16.8	245	1 ERB2_MOUSE	P70424 mus musculu
18	727	10.7	1363	1 ILPR_BRALA	O02466 branchiosto
19	699	10.3	1300	1 IRR_MOUSE	Q9wt14 mus musculu
20	696	10.2	1383	1 INSR RAT	P15127 rattus norv
21	695.5	10.2	1372	1 INSR_MOUSE	P15208 mus musculu
22	695	10.2	1382	1 INSR_HUMAN	P06213 homo sapien
23	693	10.2	1607	1 MIPR_LYMSI	Q25410 lymaea sta
24	690	10.1	1297	1 IRR_HUMAN	P14610 lymaea sta
25	685.5	10.1	1300	1 IRR_CAVPO	P14617 cavia porce
26	685	10.1	1477	1 HTK7_HYDAT	Q25197 hydra atten
27	641	9.4	1367	1 IG1R_HUMAN	P08069 homo sapien
28	623	9.1	1373	1 IG1R_MOUSE	Q60751 mus musculu
29	619.5	9.1	1370	1 IG1R RAT	P24062 rattus norv
30	618	9.1	1390	1 INSR_AEDAE	Q93105 aedes aegyp
31	615	9.0	2146	1 INSR_DROME	P09208 drosophila
32	605	8.9	987	1 EPB4_HUMAN	P54760 homo sapien
33	599.5	8.8	984	1 EPB1_CHICK	Q07494 gallus gall

34	596.5	8.8	984	1 EPB1 RAT	P09759 rattus norv
35	590.5	8.7	977	1 EPB2_MOUSE	Q03145 mus musculu
36	590.5	8.7	984	1 EPB1_HUMAN	P54762 homo sapien
37	588	8.6	902	1 EPBB_XENLA	Q91736 xenopus lae
38	588	8.6	1114	1 RET_HUMAN	P07949 homo sapien
39	585	8.6	987	1 EPB4_MOUSE	P54761 mus musculu
40	584.5	8.6	976	1 EPB4_MOUSE	P29317 homo sapien
41	580.5	8.5	985	1 EPB4_XENLA	Q91571 xenopus lae
42	573.5	8.4	1053	1 FAK1_CHICK	Q00944 gallus gall
43	569	8.4	757	1 HT16_HYDAT	P53356 hydra atten
44	569	8.4	1068	1 FAK1_XENLA	Q91738 xenopus lae
45	563	8.3	1052	1 FAK1_MOUSE	P34152 mus musculu

#### ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell surface receptor HER2) (MLN 19).  
GN ERBB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N., Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erbB-2 gene to epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A., McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J., Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
CC -I- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX, ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-ALPHA AND AMPHIREGULIN.  
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
CC -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).  
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.





```
FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 87.1%; Score 5937; DB 1; Length 1257;
Best Local Similarity 87.0%; Pred. No. 5.7e-308;
Matches 1094; Conservative 52; Mismatches 109; Indels 2; Gaps 2;

QY 1 MELAALCRWGLLLALLPPGAASTVCTGTDMLRLPASPTHLDMLRLHYQGCQVVGML 60
DB 1 MELAAWCRWGFLLALLPPGIAGTVCVTGTDMLRLPASPTHLDMLRLHYQGCQVVGML 60
QY 61 ELYLPTNASLSFLQDIOEVQGVLIHQNQVQPLQRLIRVGTQOLFEDNYALAVLDNG 120
DB 61 ELYTVFANASLSFLQDIOEVQGMIAHQNVRVQPLQRLIRVGTQOLFEDNYALAVLDNR 120
QY 121 DPLNNTTPVT-GASPGGLRELQRLSLEILKGVLIQRLNPOLCYQDQTLWKDIFHKNQOL 179
DB 121 DQDQNVAASTPGRTPEGLRELQRLSLEILKGVLIQRLNPOLCYQDQVLMWLVKQVKNQOL 180
QY 180 ALTLDTNRSRACHPCSPMKSCRSWGSESDCQSLTRTVACGGACRCKGPLPTDCCHQ 239
DB 181 APVDITNRSRACPCAPACKDNHCWGESPEDCQLTGTCISGCARCKGRPLPTDCCHQ 240
QY 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMPNPEGRYTFGASCVTAC 299
DB 241 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMHNPEGRYTFGASCVTTC 300
QY 300 PYNLSLTDVGSCTLVCPHLNQVBTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKFIGI 359
DB 301 PYNLSLTDVGSCTLVCPPPNNQVETAEDGTQRCCKSKPCARVCYGLGMEHLRGARITSD 360
QY 360 TELEFAGCKKIFGSLAFLPESDGPASNTAPLOEQLOVFTLEITGYLIISAWPDSL 419
DB 361 NVQEPDGGCKKIFGSLAFLPESDGPSSGIAFLRPEQLQVFTLEITGYLIISAWPDSL 420
QY 420 PDLVSFQNLQVIRGRILHNGAYSLTLQGLGISWGLRSLRELGSGLALIHNNTHLCFVHT 479
DB 421 RDLVSFQNLRIIRGRILHDGAYSLTLQGLGIHSLGRSLRELGSGLALIHNAHLCFVHT 480
QY 480 VPWDQLFRNPQALLHTANRPEDE-CVGEGLACHOLCARGHCWGPGTQVCNCSQFLRQ 538
DB 481 VPWDQLFRNPQALLHSGNRPEEDLCVSSGLVCNSLCAHGHCHWGPGTQVCNCSHFLRQ 540
QY 539 ECVEECRVLOGLPREYVNAHCLPCHPECPONGSVTCFPGPADOCVACHYKDPFCVA 598
DB 541 ECVEECRVWKGLPREYVSDKRLPCHPECPONGSVETCFGSEADQCAACHYKDSSCVA 600
QY 599 RCPGSKVPLSYMPIWKFPDEGACQPCPINTHSCVDLDDKGCPEAQRASPLTISVSAV 658
DB 601 RCPGSKVPLSYMPIWKFPDEGICQPCPINTHSCVDLDERGCPAEQASPTVITAV 660
QY 659 VGILLVVLGVVFGILIKRQKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRLKET 718
DB 661 VGVLFLILVVLGVVILIKRQKIRKYTMRLQLQETELVEPLTPSGAMPNQAMRLKET 720
QY 719 ELRKVKVLGSGAFGYVGIWIPDGENVKIPVAIKVLRNTSPKANKEILDAYMAGVG 778
DB 721 ELRKVKVLGSGAFGYVGIWIPDGENVKIPVAIKVLRNTSPKANKEILDAYMAGVG 780
QY 779 SPVSRLLGICLTSTVQLVTLMPYGCLLDHHVRENRGRGLGSQDLLNWCWQIAKGSYLE 838
```

```
DB 781 SPVSRLLGICLTSTVQLVTLMPYGCLLDHHVRENRGRGLGSQDLLNWCWQIAKGSYLE 840
QY 839 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESILRRR 898
DB 841 VRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGKVPKIMMALESILRRR 900
QY 899 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPICTIDVYMIWVC 958
DB 901 FTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPQPPICTIDVYMIWVC 960
QY 959 WMIIDSECRPRFRELVSERFMRARDQRFVVIQNEDLGPASPLDSTFYRSLLEDGMDLV 1018
DB 961 WMIIDSECRPRFRELVSERFMRARDQRFVVIQNEDLGPSSPNDSTFYRSLLEDGMDLV 1020
QY 1019 DAEYLVFQGGFCFDPDAPGAGMVHHRHSSTRSGGDLTLGLEPSEEEAPRSLAPS 1078
DB 1021 DAEYLVFQGGFFSDPTPGTSTAHRHRSSSTRSGGELTLGLEPSEEGPRSLAPS 1080
QY 1079 EGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPQPE 1138
DB 1081 EGAGSDVFDGLAMGVTKGQSLSPHLSPLQRYSEDTPLPPLPETDGYVAPLACSPQPE 1140
QY 1139 VYNQPDVVRQPPSPREGPLPAARPAAGATLERAKTILSPGKGVVQKVFAGGAVENPEYL 1198
DB 1141 VYNQSEVQPPPLTPEGLPVPVRPAGATLERKTLSPGKGVVQKVFAGGAVENPEYL 1200
QY 1199 FQGGAAPOPHPPAFSPAFDNLVYWDQPPPPGAPPSTFKGTPTAENPEYLGLDV 1255
DB 1201 PREGTASPPHSPAFSPAFDNLVYWDQNSSEGGPPSPNEGTPGTAENPEYLGLDV 1257

RESULT 3
ERB2_MESAU
ID ERB2_MESAU STANDARD; PRT; 1254 AA.
AC Q60553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; PubMed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255(1994)
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC  
CC  
CC  
DR EMBL: D16295; BAA03801.1; -  
DR HSSP: P11362; 1FGK.  
DR InterPro: IPR000494; EGPR\_L domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR004019; YLP\_motif.  
DR Pfam: PF00069; pkinase; 1.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR Pfam: PF02757; YLP; 2.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; TyrKG; 1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Proto-oncogene; Disease mutation.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 653 675 POTENTIAL.  
FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 158 368 CYS-RICH.  
FT DOMAIN 472 644 CYS-RICH.  
FT DOMAIN 720 987 PROTEIN KINASE.  
FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
FT BINDING 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.  
FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 567 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21P2B81 CRC64;  
Query Match 86.9%; Score 5921.5; DB 1; Length 1254;  
Best Local Similarity 86.6%; Pred. No. 3.8e-307;  
Matches 1087; Conservative 61; Mismatches 106; Indels 1; Gaps 1;  
Qy 1 MELAAACRWGLLLALLPFGAASSTOVCCTGDMKRLRLPASPEHLDMLRHLGYCCQVQGNL 60  
Db  
1 MELAAACRWGLLLALLSFGASCTOVCCTGDMKRLRLPASPEHLDIVRHLGYCCQVQGNL 60

Qy 61 ELTYLPTNASLSFLQDIOEVQGVYVLIHQAHQVQVPLQRLIRIVRGTLQFEDNYALAVLDNG 120  
Db  
61 ELTYLPANATLSFLQDIOEVQGVYVLIHQAHQVQVPLQRLIRIVRGTLQFEDKYALAVLDNR 120  
Qy 121 DPLNNTPTVTGASPGGLRELOLRSLTEILKGGVLIQORNQOLCYQDTILKDKIIFHKNOLA 180  
Db  
121 DPLDNVTTATGRTPEGLRELQRLSLEILKGGVLIQORNQOLCYQDTILKDKIIFHKNOLA 180  
Qy 181 LTLIDTNRSRACHPCSMCKGSRGSESSDQSLTRTVTCAGGCARCKGKPLPTDCCHEQC 240  
Db  
181 PVDIDTNRSRACPPCAPACKDNHCWGPEDCQTLTGTTAPRAVPAARARLPTDCCHEQC 240  
Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Db  
241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 300  
Qy 301 YNYLSTDVGSCTLVCPLNHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 360  
Db  
301 YNYLSTDVGSCTLVCPLNHNSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACP 360  
Qy 361 ELEFAGCKKI FGS LAF L P E S F D G D P A S N T A P L Q E O L Q V F E T L E E I T G Y L Y I S A W P D S L P 420  
Db  
361 IOEFAGCKKI FGS LAF L P E S F D G N P S S G I A P L T P B O L Q V F E T L E E I T G Y L Y I S A W P D S L H 420  
Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSLRELGLALIHNTHLFCFVHTV 480  
Db  
421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISWLGRLSLRELGLALIHNTHLFCFVHTV 480  
Qy 481 PWDQLFRPHQALLHTANRPEDECYGEGLACHQLCARGHCWGPGPTQCVCNCSOFLRGQEC 540  
Db  
481 PWDQLFRPHQALLHSGNPSEEEGLKDFACYPLCAHGHGWCWPGPTQCVCNCSHFLRGQEC 540  
Qy 541 VEECRVLQGLPREYVNAHRLCPHPECPQNGSVTCFGPEADQCACAHYKPPFCVCARC 600  
Db  
541 VKECNVWGLPREYVNGHRLCPHPECPQNGSVTCFGPEADQCACAHYKPPFCVCARC 600  
Qy 601 PSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG 660  
Db  
601 PSGVKPDLSYMPIMKVPDEEGWCQPCINCTHSCVDLDDKGPAPAEQASPLTSIVSAVVG 660  
Qy 661 ILLVVVLGVVFGILIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Db  
661 ILLFLVLGVVFGILIKRQOKIRKVTMRRLLOETELVEPLTPSGAMPNQAQMRILKETEL 720  
Qy 721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
Db  
721 RKVKVLGSGAGFTVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEILDEAYVMAGVGP 780  
Qy 781 YVSRLLGLCLTSTVOLVQLMPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEDVVR 840  
Db  
781 YVSRLLGLCLTSTVOLVQLMPYGLLDHVRNRRGLSGQDLLNWCQIAKMSYLEDVVR 840  
Qy 841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESILRRRFT 900  
Db  
841 LVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKIMWALESILRRRFT 900  
Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVTMVMVKCM 960  
Db  
901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPPOPICTIDVTMVMVKCM 960  
Qy 961 IDSECRPRFRELVSFSESMARDPQRVVIQNEIDLGPASPLDSTFYRSLLDMDMDGLVDA 1020  
Db  
961 IDSECRPRFRELVSFSESMARDPQRVVIQNEIDLGPASPLDSTFYRSLLDMDMDGLVDA 1020  
Qy 1021 EYLVPOQGFCTPDPAPGAGMVHRRSSSTRSGGDLTLGLBFSSEEAAPRSLAPSEG 1080  
Db  
1021 EYLVPOQGFCTPDPAPGAGMVHRRSSSTRSGGDLTLGLBFSSEEAAPRSLAPSEG 1080  
Qy 1081 AGSDVFDGDLGMAKGLQSLPTHPDPSPLQRYSEDPVPLPSETGYVAPLTCSPQEVY 1140  
Db  
1081 AGSDVFEGLGMAKGLQSPQLSPLQRYSEDPVPLPSETGYVAPLTCSPQEVY 1140

QY 1141 NQPDVVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAGGAVENPEYLTPQ 1200  
 DB 1141 NQPEVRPQLTPEGFLPPVRPAGATLERPKTLSPGKNGVVKDVFAGGAVENPEYLVPR 1200  
 QY 1201 GGAAPQPPPPAFSPAFDNLVYWDQDPPRGPAPPSTFKGTPTAENPEYILGLDVPV 1255  
 DB 1201 GGSASQPH-PPALCPAFDNLVYWDQDPPSERGSPNPTFEGTPTAENPEYILGLDVPV 1254

RESULT 4  
 EGFR\_HUMAN STANDARD; PRT: 1210 AA.  
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMDS; Q9UMG5; Q92795; Q00732;  
 AC Q00688; Q9BZS2; Q9BZC9; Q9GZX1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 DE protein-tyrosine kinase Erbb-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RL Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=95382957; PubMed=7654368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reiter J.L., Maible N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RC TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schell Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Maible N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schell C.M.,  
 RA Lampland A.L., Balasubramanian S., Crossley T.O., Magnuson T.R.,

RA Maible N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor.";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarek L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";  
 RL Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells.";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RL Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RL Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]



```

QY 1141 NOPDVRPQPPSPREGPLPAARAGATLERAKTLSPGKNGVVKDVFAFGGAVENPEYL-TP 1199
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1094 NO-SVPRKPGAGSVQNPVHNQPLNP-----APSRDPHYQD--PHSTAVGNPEVLNTV 1142
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1200 QGGAAPQPHPPAFSPAFNLYWDQ-----DP-----PERGAPPSTFKGTPTA 1243
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1143 Q-----PTCVNSTFDSPAHWAQKGSHQISLDNPDYQDFFPKPEAKPNGIFKGS-TA 1192
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1244 ENPEYL 1249
    ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1193 ENAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD; PRT; 1210 AA.
AC Q01279;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (BC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.;
RT "Promoter region of the murine fibroblast growth factor receptor 2
   (bek/KGFR) gene.";
RL Oncogene 7:1957-1962(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
   in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;
RA Hibbs M.L.;
RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=8125255;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,
RA Jenkins N.A., Lee D.C.;
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
   receptor tyrosine kinase.";
RL Genes Dev. 8:399-413(1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morsee B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
   binding site.";
RL Oncogene 6:673-676(1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Eisenger D.P., Serrero G.;
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC tyrosine phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78987; CAA55587.1; -
CC EMBL; U03425; AAA17899.1; -
CC EMBL; X59698; CAA42219.1; -
CC EMBL; L06864; AAA53029.1; -
CC EMBL; Z12608; CAA78249.1; -
CC HSSP; P11362; 1FGK.
CC MGD; MGI:95294; Egfr.
CC InterPro; IPR000494; EGFR_L domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recep_L domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; Tyrc; 1.
CC PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC PROSITE; PS5011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 1210 EPIDERMAL GROWTH FACTOR RECEPTOR.
CC DOMAIN 25 647 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 648 670 POTENTIAL.
CC DOMAIN 671 1210 CYTOPLASMIC (POTENTIAL).
CC REPEAT 75 300 APPROXIMATE.
CC REPEAT 390 600 APPROXIMATE.
CC DOMAIN 1028 1071 SER-RICH.
CC DOMAIN 714 981 PROTEIN KINASE.
CC NP_BIND 720 728 ATP (BY SIMILARITY).
CC BINDING 747 747 ATP (BY SIMILARITY).
CC ACT_SITE 839 839 BY SIMILARITY.
CC DISULFID 130 199 BY SIMILARITY.
CC DISULFID 194 207 BY SIMILARITY.
CC DISULFID 215 223 BY SIMILARITY.
CC DISULFID 219 231 BY SIMILARITY.
CC DISULFID 232 240 BY SIMILARITY.
CC DISULFID 236 248 BY SIMILARITY.
CC DISULFID 251 260 BY SIMILARITY.
CC DISULFID 264 291 BY SIMILARITY.
CC DISULFID 295 307 BY SIMILARITY.
CC DISULFID 311 326 BY SIMILARITY.
CC DISULFID 329 333 BY SIMILARITY.
CC DISULFID 506 515 BY SIMILARITY.
CC DISULFID 510 523 BY SIMILARITY.
CC DISULFID 526 535 BY SIMILARITY.
CC DISULFID 539 555 BY SIMILARITY.
CC DISULFID 558 571 BY SIMILARITY.
CC DISULFID 562 579 BY SIMILARITY.
CC DISULFID 582 591 BY SIMILARITY.
CC DISULFID 595 617 BY SIMILARITY.
CC DISULFID 620 628 BY SIMILARITY.
CC DISULFID 634 636 BY SIMILARITY.
CC MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
CC MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

```









```
Db      -536  CNLYDGEFEFENGSIVCEDPOCEKQWEDGLLTCHGPGPDNCTKCSHFDPNCVKECPD 595
Qy      603  GUKPDLNYPINWIKFPDEBEGACQPCPNCTHSCVDLDDKGC-----PAEQRASPL 651
Db      596  GLQGANSF--IFKYADPDRECHPCPNCTQGCNGPTSHDCIYYPWTGHSTLPOHAR-TFL 652
Qy      652  TSIVSAVW-GILLVVLGVVFGILLKRRROOKIRKYMRLLOETELVELPLTSGAMPNQA 710
Db      653  --IAGVIGGLFILVIVGUTFAVYVRKSIK-KRALRRFL-ETELVEFLTSGGTAPNA 708
Qy      711  QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTPSKANKEILDE 770
Db      709  QLRILKETELRKVKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRNTPSKANKEILDE 768
Qy      771  AVYVAGVSPYSRLLIGLITSTVQLVTQIMPYGCLLDHVRNRLGSLQDILLNWCMTA 830
Db      769  ALIMASMDPHLVRLLGVLCSPTIQLVTQIMPYGCLLDHVRNRLGSLQDILLNWCMTA 828
Qy      831  KGMVLYEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDDDEYHADGGKVPKWA 890
Db      829  KGMVLYEDVRLVHRDLAARNVLKSPNVHKITDFGLARLLDDDEYHADGGKVPKWA 888
Qy      891  LESILRRRTHOSDVMYSYGVYVWELMTFGAKPYDGPAREIPDLLEKGERLPQPPICTID 950
Db      889  LECIHYRKFTHOSDVMYSYGVYVWELMTFGKPYDGPITREIPDLLEKGERLPQPPICTID 948
Qy      951  VYVWVVKWMDISECRPRFELVSESRMARPOQFVWIONED-LGPASPLDSTFVRSLL 1009
Db      949  VYVWVVKWMDISADSRPKFELAAEFRRMARPOQFVWIONED-LGPASPLDSTFVRSLL 1008
Qy      1010  EDDMDGLVDAAEYLVPQGFPCDPAPCAGAGVHRRHRSSTRSGGDLTLGLLEPSEEE 1069
Db      1009  DEEDLMDAAEYLVP-QAFNIPP-----IYTSRARDNRS-----EIGHSPPPAY 1056
Qy      1070  APRS-----PLAP-SEGAGSDVDFDGLGMAAGKLOS 1100
Db      1057  TMSGNQFYVYRGGFAAEQGVSVYRAPTSITPEAPVAGATAEIFDDSCCNGTLRKPA 1116
Qy      1101  LPTBPSPLORSEDTVPPLPS-----ETDGYVAPLTCSPQPEYVQPDVVRPQPPSR 1153
Db      1117  PHVQDSSTQRYSDPTVPAPERSPRGELDEEGYMTMRDKPKQEYLNVE----- 1167
Qy      1154  EGPLPAARPAAGATLERAKTSLPGKGVKVDVAFGAVENPEYLTPOGGAAPQPHPPPA- 1212
Db      1168  ENPFVSR-----KNGDLQ-----ALDNPETHNASNG-----PPKAE 1199
Qy      1213  -----FSPAFDNLVYWDQDPPERGA--PPSTF 1237
Db      1200  DEYVNEPLYNTFANTLGRKAEYLNKILSMPEKAKAFDPDYNWHSUPRSTLQHPDYL 1259
Qy      1238  KGTPT-----AENPEYL 1249
Db      1260  QEYSTKYFYKQNGRIRPIVAENPEYL 1285

RESULT 7
ERR4_RAT
ID  ERB4_RAT  STANDARD;  PRT;  1308 AA.
AC  Q62956; Q922N7;
DT  15-DEC-1998 (Rel. 37, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN  ERB4 OR TYRO-2.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]_TaxID=10116;
RP  SEQUENCE FROM N.A.
RC  TISSUE=Heart;
RX  MEDLINE=98221155; PubMed=9553078;
RA  Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
```

```
RA  Marchionni M.A., Kelly R.A.;
RT  "Neuregulins promote survival and growth of cardiac myocytes.
RT  Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT  ventricular myocytes.";
RL  J. Biol. Chem. 273:10261-10269(1998).
RN  [2]
RP  SEQUENCE OF 848-901 FROM N.A.
RC  TISSUE=Sciatic nerve;
RX  MEDLINE=91222560; PubMed=2025425;
RA  Lai C., Lemke G.;
RT  "An extended family of protein-tyrosine kinase genes differentially
RT  expressed in the vertebrate nervous system.";
RL  Neuron 6:691-704(1991).
RN  [3]
RP  SEQUENCE OF 1031-1198 FROM N.A.
RC  STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX  MEDLINE=97184212; PubMed=9030624;
RA  Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;
RT  "Expression of neuregulins and their putative receptors, ErbB2 and
RT  ErbB3, is induced during Wallerian degeneration.";
RL  J. Neurosci. 17:1642-1659(1997).
CC  -!- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC  2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC  NTK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC  NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC  -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC  tyrosine phosphate.
CC  -!- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC  RECEPTORS (POTENTIAL).
CC  -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC  NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC  OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC  RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC  HEART.
CC  -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC  RESIDUES (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AF041838; AAD08899.1; --
CC  EMBL; U52531; AAC53051.1; --
CC  HSP; P11362; IFGK.
CC  InterPro; IPR000494; EGFR_L domain.
CC  InterPro; IPR000719; Euk_pkinase.
CC  InterPro; IPR002174; Furin-like.
CC  InterPro; IPR001245; Tyr_pkinase.
CC  InterPro; IPR004019; YLP_motif.
CC  Pfam; PF00757; Furin-like; 1.
CC  Pfam; PF00069; pkinase; 1.
CC  Pfam; PF01030; Recep_L_domain; 2.
CC  Pfam; PF02757; YLP; 2.
CC  PRINTS; PR00109; TYRKINASE.
CC  ProDom; PD000001; Euk_pkinase; 1.
CC  SMART; SM00261; FU; 4.
CC  SMART; SM00219; TyrKc; 1.
CC  PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC  PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC  PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW  Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW  Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT  SIGNAL 1 25
FT  CHAIN 26 1308
FT  DOMAIN 26 651
FT  TRANSMEM 652 675
FT  DOMAIN 676 1308
FT  CYTOPLASMIC (POTENTIAL).
```



01-JAN-1990	(Rel. 13, Created)
16-OCT-2001	(Rel. 40, Last sequence update)
15-JUN-2002	(Rel. 41, Last annotation update)
Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).	
GN	XMRK OR TU.
OS	Xiphophorus maculatus (Southern platyfish).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC	Acanthopterygii; Acanthopterygii; Percomorpha; Atherinomorpha;
OC	Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX	NCBI_TaxID=8083;
OX	[1]
RN	SEQUENCE FROM N.A.
RP	MEDLINE=90015140; PubMed=2797166;
RA	Wittbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,
RA	Telling A., Robertson S.M., Scharlt M.;
RA	"Novel putative receptor tyrosine kinase encoded by the melanoma-
RT	inducing Tu locus in Xiphophorus.";
RL	Nature 341:415-421(1989).
[2]	
RP	REVISION TO 515.
RA	Scharlt M.;
RL	Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC	-!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
CC	-!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC	tyrosine phosphate.
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
CC	-!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
EMBL	X16891; CAA34770.2; --
PIR	S06142; S06142.
HSP	P11362; IFGK.
DR	InterPro: IPR000494; EGFR_L domain.
DR	InterPro: IPR000719; Euk_pkinase.
DR	InterPro: IPR002174; Furin-like.
DR	InterPro: IPR002290; Ser_thr_pkinase.
DR	InterPro: IPR001245; Tyr_pkinase.
DR	Pfam: PF00069; pkinase; 1.
DR	Pfam: PF00757; Furin-like; 1.
DR	Pfam: PF01030; Recep_L domain; 2.
DR	PRINTS: PR00109; TYRKINASE.
DR	ProDom: PD000001; Euk_pkinase; 1.
DR	SMART: SM00261; FU_5.
DR	SMART: SM00220; S_TKC; 1.
DR	SMART: SM00219; TYRK; 1.
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW	Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW	Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
FT	SIGNAL
FT	1 25
FT	CHAIN 26 1167
FT	DOMAIN 26 642
FT	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM 643 665
FT	POTENTIAL.
FT	DOMAIN 666 1167
FT	CYTOPLASMIC (POTENTIAL).
FT	DOMAIN 710 977
FT	NP BIND 716 724
FT	BINDING 743 743
FT	ACT SITE 835 835
FT	BY SIMILARITY.
FT	DISULFID 195 204
FT	BY SIMILARITY.
FT	DISULFID 199 212
FT	BY SIMILARITY.
FT	DISULFID 220 228
FT	BY SIMILARITY.
FT	DISULFID 224 236
FT	BY SIMILARITY.
FT	MELANOMA RECEPTOR PROTEIN-TYROSINE
FT	KINASE.

```
Db 590 IPRCPHGLGDDFL-TWKYADKNGQCPCHQNCQTQCSCGRLSGCRGD-IVSHSLAVG 647
Qy 657 AVUGILLVVLGVVFGILIKRQOKIRKYTRRLLQETELVEPLTPSGAMPNQAMRLK 716
Db 648 LVSGULLTIVALLIVLLRRRRIK-RKRTIRCLLOEKELVEPLTPSQAPNQAFRLIK 706
Qy 717 ETELKRVKVLGSGAGFYVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKEILDYAYNAG 776
Db 707 ETEFKDORVLGSGAGFYVYKGLMNPGENIRIPVAIKVLRNTPSPKANKEILDYAYNAG 766
Qy 777 VGSFVYVRLGICLTSTVQLVTLQMPYCGCLLDVYRNRGRGLSGDILNWCQIAKGSYL 836
Db 767 VDRPHVCRLLGICLTSAVLVTLQMPYCGCLLDVYRNRGRGLSGDILNWCQIAKGSYL 826
Qy 837 EDVRLVHRDLAARNVLKPNHVKITDFGLARLLDIDETEHYADGKVPKIKWMALESILR 896
Db 827 EERHLVHRDLAARNVLKPNHVKITDFGLSKLLTADKEKYQADGGKVPKIKWMALESILQ 886
Qy 897 RRFTHQSDVWSYGVTVVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPICITIDYIMV 956
Db 887 WTYTHQSDVWSYGVTVVWELMTFGSKPYDGIPIAREIPDLLEKGERLPQPPICITIEVYMI 946
Qy 957 KCMWIDSECPREFELVSESRWARDQRFVVIQNEIDLGPASPLDSTFYRSLLEDDEMDGD 1016
Db 947 KCMWIDSPRPRFRELVEFSQWARDPSRYLVIOG---NLPSLSDRLFSRLSSDD--D 1001
Qy 1017 LVDAEYLVPOOGFCFDPAPGAGGMVHRRHSSTRSGGDLTLGLEPSEEAAPRPLA 1076
Db 1002 VVDAEYLVPKRI-----NRQGS-----EPCI 1024
Qy 1077 PSEGAGSDVFDGLGMAAGLQSLPDPSPLOQYSEDPTV-PLPSETDGYVAPLTCSP 1135
Db 1025 PPTGH-----PVRENSITLRNISDPTQNALEKDLGDH----- 1056
Qy 1136 QPEVNPDPVRPQP-----PSPRE-----GPLP-AARPAGATLERAKTLSPKNGVVK 1182
Db 1057 --EVNPGGETSRRLSDIYNPTVEDTDCGVPVLSLSQEAETNFSPEYLTNTQNSL-- 1112
Qy 1183 DVFAFGAVENPEYLTQGGAAPOPHPPAFSPAFDNLVYWDQPPPERGAPPSTFKGTPT 1242
Db 1113 -PLVSSGSMDDPDY---QAG-----YQAFAF-----LPQTGALTNGMGFLPA 1149
Qy 1243 AENPEYLG 1250
Db 1150 AENLEYLG 1157

RESULT 9
ERB3_HUMAN
ID ERB3_HUMAN STANDARD; PRT; 1342 AA.
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
```

```
RA Plowman G.D., Whittier G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX TISSUE=Placenta;
RX MEDLINE=9328282; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase.";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -!- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -!- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M29366; AAA35790.1; --
CC EMBL; M34309; AAA35979.1; --
CC EMBL; S61953; AAB26935.1; --
CC PIR; A36223; A36223.
CC HSSP; P11362; IFGK.
CC Genew; HGNC:3431; ERBB3.
CC MIM; 190151; --
CC InterPro; IPR000494; EGFR_L_domain.
CC InterPro; IPR000719; Euk_pkinase.
CC InterPro; IPR002174; Furin-like.
CC InterPro; IPR001245; Tyr_pkinase.
CC Pfam; PF00069; pkinase; 1.
CC Pfam; PF00757; Furin-like; 1.
CC Pfam; PF01030; Recept_L_domain; 2.
CC ProDom; PD000001; Euk_pkinase; 1.
CC SMART; SM00261; FU; 3.
CC SMART; SM00219; TyrKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
CC PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
CC PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
CC Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
CC Alternative splicing.
FT SIGNAL 1 19
FT CHAIN 20 1342
FT DOMAIN 20 643
FT TRANSMEM 644 664
FT DOMAIN 665 1342
FT DOMAIN 709 966
FT NP_BIND 715 723
FT BINDING 742 742
FT ACT_SITE 834 834
FT DISULFID 186 194
FT DISULFID 190 202
```

```
FT DISULFID 210 BY SIMILARITY.
FT DISULFID 214 BY SIMILARITY.
FT DISULFID 227 BY SIMILARITY.
FT DISULFID 231 BY SIMILARITY.
FT DISULFID 243 BY SIMILARITY.
FT DISULFID 246 BY SIMILARITY.
FT DISULFID 255 BY SIMILARITY.
FT DISULFID 259 BY SIMILARITY.
FT DISULFID 290 BY SIMILARITY.
FT DISULFID 301 BY SIMILARITY.
FT DISULFID 323 BY SIMILARITY.
FT DISULFID 327 BY SIMILARITY.
FT DISULFID 500 BY SIMILARITY.
FT DISULFID 509 BY SIMILARITY.
FT DISULFID 517 BY SIMILARITY.
FT DISULFID 520 BY SIMILARITY.
FT DISULFID 533 BY SIMILARITY.
FT DISULFID 552 BY SIMILARITY.
FT DISULFID 556 BY SIMILARITY.
FT DISULFID 573 BY SIMILARITY.
FT DISULFID 576 BY SIMILARITY.
FT DISULFID 589 BY SIMILARITY.
FT DISULFID 610 BY SIMILARITY.
FT DISULFID 613 BY SIMILARITY.
FT DISULFID 621 BY SIMILARITY.
FT DISULFID 629 BY SIMILARITY.
FT CARBOHYD 126 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 250 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 408 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 414 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 437 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 469 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 522 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 566 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 616 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 141 EILSGGVIEKNDKLCMDTIDWRDVIWRDRDAEIVVKNR
SC -> GQFMPSPGLTPQADQWVLLDDPRLTLTSSSK
VPTLAIV (IN SHORT ISOFORM).
MISSING (IN SHORT ISOFORM).
E -> G (IN REF. 2).
E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
```

Query Match 35.6%; Score 2423.5; DB 1; Length 1342;  
Best Local Similarity 40.5%; Pred. No. 2.2e-121;  
Matches 532; Conservative 195; Mismatches 452; Indels 135; Gaps 34;

```
QY 10 GLLALLPGAA--STQVCTGTDMLKRLPASBETHLDMRLHLYQGQVQVQGNLELYLPT 67
DB 11 GLLFSLARGSEVGNQAVCPGLTNGLSVTGDAENQYTLKLYRCEVVMGNLEIVLIGH 70
QY 68 NASLSFLODIQEVQVGLIAHNOVQVPLQRLIRIVRGTOLEFEDNVALAVLDNGDPLNNTT 127
DB 71 NADLSFLQWIREVTGVVLVAMNEFTLPLNLRVVRGTQVYDQKFAIFVW-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSLTEILKGGVLIQORNPLQCYQDTILWKDIFHKNNQALTLIDTN 187
DB 126 ---NSSHALRQLTLQTLTEILSGGVYIEKNDKLCMDTIDWRDVIWRDRD--AEIVVKD 178
QY 188 RSRACHPCSPMCKSRCSGESSEDCOSLTRTVCAGGC-ARCKGPLPTDCCHQCAAGCTG 246
DB 179 NGRSCPPCHEVCKG-RCWGPGEDECOTLTKTICAPQCNGHCFCGPNPQOCHDECAGGCSG 237
QY 247 PKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNPEGRYTFGSCVACPXYNLSLT 306
DB 238 PQDTCFACRHENDSGACVPRCPQPLVYNKLTFOLEPNHTKYQYGGVCVASCNPNFV-V 296
QY 307 DVGSCTVCLPHNQEVTAEDGTQRCCKSKPCARVCYGLGMOYIKANSKF--IGITELE- 363
DB 297 DQTSVCVRACPPDKMEVD-KNGLKMCPECPGGLCPKACEGTG-----SGSRFQVDSNIDG 350
QY 364 FAGCKKIFGSLAFLESFDGDPASNTAPLOLOVQFETLEITGYLVIASWPSLPLDLS 423
DB 351 FVNCTKIIGNLDFLTGLNGDPPHKKIPALDPEKLVNFRVREITGYLNIQSWPPHMFNS 410
QY 424 VFQNLQVIRGRILHNGAYS-LTQGLIGISWLGRSLRELGSGLALIHNNTHLCFVHTVPW 482
DB 411 VFSNLTITIGRSLYNRGFSLLIMKLNVTSLGFRSLKEISAGRIYISANRQCLCYHHSLNW 470
```

```
QY 483 DQLFRNPHOALLHTA-NRPEDEBCVGEGLACHQLCARGHCMGPQPTQCVCNSQFLRGQCV 541
DB 471 TKVLRGPTTEERLDIKHNRPRRDCAEGKVCDCPLCSSGGCMGPGQCLSCRNYSRGVCV 530
QY 542 EECRVLOGLPREYVNAHCLCPHCEQOPQNGSVTCFGEADQCVACAHYKDPFCVACP 601
DB 531 THCNFLNGEPREFAEAEFCFCHPECPMEGTATCNGSGSDTCAQCAHFRDGHCVSSCP 590
QY 602 SGVKPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPAEQRA---SPLTSTVSA 657
DB 591 HGVLG--AKGPIYKYPDVQNECRPCHENTCQCKGPELOCLQOTLVLIQKTHLTALTV 648
QY 658 VVGILLVVLGVVFGILIKRQOKIR-KYTMRLLOETELVELPTPSGAMPNOAQMRLK 716
DB 649 IAG--LVWIFMMLGGTFLYWRGRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIFK 705
QY 717 ETELKVKVKGSGAGCTVYKGIWIPQGENVKIPVAIKVIRENTSPRANKELIDEAYVMAG 776
DB 706 ETELKRLKVLGSGVFGTVHKGVMIPEGESIKVIEDKSGROSFOAVTDHMLAIGS 765
QY 777 VGSPYVSRLLGCLTSTVOLTPQMPYGLLDHVRNRRGLSGQDLLNMCQIAKMSYL 836
DB 766 LDHAHIVRLGLGCPGSSLOLVQYLPGLSLLDHRVQRHAGLGPQLLLNHWQVQIAKMYL 825
QY 837 EDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMMALESILR 896
DB 826 EEBGMVHRNLAARNVLLKSPSQVQVADFGVADLLPPDDKQLLYSEAKTPIKMMALESIFH 885
QY 897 RRFTHQSDVMSGVTVWELMTGAKYDGIIPAREIPDLEKGERLPOPPICTDIVVMV 956
DB 886 GKYTHQSDVMSGVTVWELMTGAEYAGYLAELAEVDPDLLEKGERLAQPOQICTDIVVMV 945
QY 957 KCWMDISECRPRELVSFEFSRMARDPQRFVIONEIDLGA---SPLDSTFYRSLLEDD 1013
DB 946 KCWMDENIRPTFKELANEFTRMARDPPRYLVIKRES-GPGIAGPEPHGLTNNKLEVE 1004
QY 1014 MGLDVAEEYLVPOQGFCCPDPAAGAGMVHRRSSSTRSGGDLTLGLPE-SEEAAPR 1072
DB 1005 LEPELDLDLEAEED-----NLTTLGSALSPLVGLNRRPGSQ 1045
QY 1073 SPLAPSEAGSDVFDGLGMAAKGLQSLPTH-D-PSPLORYSEDTVPPLD-----SETD 1125
DB 1046 SLLSPSSGY-MPMNQNLGESCQESAVSGSSERCPRVSLH-----PMPRGCLASESE 1098
QY 1126 GYVA-----PLTCSPOPE-----YVNPQDVRPQPPSPREC----- 1156
DB 1099 GHVTGSEAELOEKVSMCRSRSRSPRGRDSAYHSQRHSLLTPTVPLSPGLEEDVNG 1158
QY 1157 --LPAARPAAGATLERAKTLSP-GKNGV-----KDVFAFGGAVENPEYLTPOGGAAPQ 1207
DB 1159 YVMPDTHLKGTPSSREGTILSSVGLSVLGTEDDED-----EYEVYNNRRRRHSP-P 1209
QY 1208 HPPAPSPAFDNLVYWD-----QDPPERGAPPSTFKGTPTAENPEYL 1249
DB 1210 HPPRPSLEELGYEYMDVGDLSASLGSTQSCPLHPVPIMPTAGTTTDEDEYEM 1263
RESULT 10
ERR3_RAT
ID ERR3_RAT STANDARD; PRT; 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
```

RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
RX MEDLINE=96096335; PubMed=8522190;  
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Kolland J.G.;  
RT "Cloning of the rat ErbB3 cDNA and characterization of the  
recombinant protein.";  
RL Gene 165:279-284(1995).  
RN [2]  
RP REVISIONS TO 85; 513 AND 565.  
RA Hellyer N.J., Kolland J.G.;  
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 922-1097 FROM N.A.  
RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
RX MEDLINE=97184212; PubMed=9030624;  
RA Carroll S.L., Miller M.L., Frohnert P.W., Kim S.S., Corbett J.A.;  
RT "Expression of neuroligins and their putative receptors, ErbB2 and  
ErbB3, is induced during Wallerian degeneration.";  
RL J. Neurosci. 17:1642-1659(1997).  
CC -!- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTKA.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
(POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -!- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
CC -!- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
PHOSPHATIDYLINOSITOL 3-KINASE.  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC -----  
DR EMBL; U29339; AAC28498.2; -;  
DR EMBL; U52530; AAC53050.1; -;  
DR HSP; P11362; LFQK.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00219; TYRK; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE\_NEG.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.  
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 19  
FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.  
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 644 662 POTENTIAL.  
FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 183 259 CYS-RICH.  
FT NP\_BIND 707 964 PROTEIN KINASE.  
FT BINDING 713 721 ATP (BY SIMILARITY).  
FT ACT\_SITE 832 832 BY SIMILARITY.  
FT DISULFID 186 194 BY SIMILARITY.  
FT DISULFID 190 202 BY SIMILARITY.  
FT DISULFID 210 218 BY SIMILARITY.  
FT DISULFID 214 226 BY SIMILARITY.  
FT DISULFID 227 235 BY SIMILARITY.

FT DISULFID 231 243 BY SIMILARITY.  
FT DISULFID 246 255 BY SIMILARITY.  
FT DISULFID 259 286 BY SIMILARITY.  
FT DISULFID 290 301 BY SIMILARITY.  
FT DISULFID 305 320 BY SIMILARITY.  
FT DISULFID 323 327 BY SIMILARITY.  
FT DISULFID 500 509 BY SIMILARITY.  
FT DISULFID 504 517 BY SIMILARITY.  
FT DISULFID 520 529 BY SIMILARITY.  
FT DISULFID 533 549 BY SIMILARITY.  
FT DISULFID 556 573 BY SIMILARITY.  
FT DISULFID 576 585 BY SIMILARITY.  
FT DISULFID 589 610 BY SIMILARITY.  
FT DISULFID 613 621 BY SIMILARITY.  
FT DISULFID 617 629 BY SIMILARITY.  
FT CARBOHYD 126 126 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 250 250 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 408 408 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 437 437 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 469 469 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 522 522 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 566 566 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 616 616 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CONFLICT 1028 1028 L -> P (IN REF. 3).  
SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDFD1E CRC64;  
  
Query Match 34.5%; Score 2351.5; DB 1; Length 1339;  
Best Local Similarity 40.9%; Pred. No. 1.4e-117;  
Matches 525; Conservative 171; Mismatches 430; Indels 159; Gaps 35;  
  
QY 3 LAALCRWGLLLALLPPGAA---STQVCTGCTDMKRLPASPETHDMLRLHLYGSCVQVQGN 59  
DB 7 LQVLC---FLSLARGSEMGNSQAVCPGTNGLSVTGDADNOYQTYLYKDYKEVVMGN 62  
QY 60 LELTYLPTNASLFLQDIOEVQGYVLIAHNOVRQVPLQRLRIVRGTLQFEDNYALAVLDN 119  
DB 63 LEIVLTGHNADLSFLQWIREVTGYVLVAMNEFSLPLNLRVVRGTVQVYDGKFAIFVM-- 120  
QY 120 GDP LNTTPTVTCASPGGLRELQRLSLTEILKGVLIOQNPOLCYQDTILWKDIFHKNNQL 179  
DB 121 ---LNYNT---NSSHALRQLKFTQLTEILSGGVYIEKNDKLMOTIDWRDVRVR-- 170  
QY 180 ALTLDTNRSRACHPCSPMKSGESSEDCOSLTRVTVCAGGC--ARKCKGLPTDCCHE 238  
DB 171 GAIVVKNNGANCPCHVECKG--RCWGPDPDCQLITKTCAPQCNRCFCGPNQCCHD 229  
QY 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTA 298  
DB 230 ECAGGCGSPQDTCFACRFNDSGACVPRCPPLVYNNKLTFFOLEPNHTKYQYGGVCVAS 289  
QY 299 CPYNYLSTDVGSCTLVCPLNHOEVTAEDGTQCEKCKSPCARVCYGL--GMQYIKANSKF 356  
DB 290 CPHNFV--VDQTFVCRACPPDKMEVD--KHGLKMCPCGGLCPKACEGTSGSRVQTVDDSN 347  
QY 357 IGITELEFAGCKKIFGSLAFLESPDGDPSANTAPLOEQLOVFEETLEITGVLYISAMP 416  
DB 348 ID----GFVNCVKILGNLDFLITGLNVDPWHKIPALDPEKLNIVRTVREITGYLNTQSWP 403  
QY 417 DSLPDLSPFQNLQVIRGRILHNGAYS--LTLOGLGISWGLSLRSLRELSGLALTHHNTLC 475  
DB 404 PHMNFVSFNSLTTIGRSLYNGFSLIMKLNVTSLGFRSLKEISAGRYVISAQQQLC 463  
QY 476 FVHTVPWDLFRNPQALLHTA--NRPEDECVGEGLACHOLCARGHCWGPGFTQVCNCSQF 534  
DB 464 YHHSNLWTRLLRGPSEERLDIKYDRPLGELAEKGVCDPLCSGGGCGWPGGQCLSCRNY 523  
QY 535 LRQGEVCEBECRYLQGLPREYVNNARHCLPCHPCQPQNGSVTCFPGPADOCVACAHYKDP 594  
DB 524 SREGVCVTNCFLOGEPRFVHEAOQFCFCHPECLPMEGTSTNGSGSDACARCAHFRDGP 583  
QY 595 FCVARCPGKVDLSYMPITWPKFPDBEGACQPCPINCTHSC--VDLDDKCPAEQASPLT 652



Db 584 HCVNSCPHILG--AKGPIYKTPDAQNECRPCHEMTCQCNCPQLQDCLGQAEVLMSKRP 641  
QY 653 SIVSAVGLLVVVLGVFGILIKRQOKIR-KYTMRLLOTELVEPLTPSGAMPNOAQ 711  
Db 642 LVIAVTVG--LAVILMILGSGFLYWRGRIQKRAMRYLGERSEIPLDPS-ERAKVYL 698  
QY 712 MRILKETLKVVLGSGAFGVYKGIWIPDGNVKIPVAIKVLRENTSPKANKILDEA 771  
Db 699 ARIFKETLRLKVLGSGVFGVHRGWIPEGESIKIPVCIKVEDKSGRQFOATDH 758  
QY 772 YVMAGVGPYSRLLICLTSTVLQVLTOLMPYGCLLDHVRENRLGSLDILNMCWQIAK 831  
Db 759 LAVGSLDHAHVRLGLCPGSSQLVQVPLGSLLDHVHQHRELPGPOLLNMGVQIAK 818  
QY 832 GMSYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETRYHADGGKVPKMMAL 891  
Db 819 GNYLLEHSMVHRDLARNVLMKSPSQVQVADGVLPPDKQLLHSEAKTPIKMMAL 878  
QY 892 ESILRRRTHQSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPCTIDV 951  
Db 879 ESIHFGKYTHQSDVMSYGVYVWELMTFGAEPYAGLRLAIPDLLEKGERLAQPICTIDV 938  
QY 952 YIMWVKMWIDSECRPRELSEFSEMRWARDPQRFVWIONEDLGPASPLDSTFYRSLLED 1011  
Db 939 YVMWVKMWIDENIRPTKELANEFTMRARDPRYLVIKRA--GPGTP--PAAEFSLVT 995  
QY 1012 DMGDLVDAEYLVPOQFFCPDPAPAGGMVHRRHSSTRSGGDLTLGLEPSEE--- 1068  
Db 996 KEL-----QEAELEPEL-----DLDLDLEAEGLA 1021  
QY 1069 -----EAPRSLAPSEG-----AGSDVFDGDLGMAKGLQSLPT 1103  
Db 1022 TSLGSALESLPTGLTRPGSQSLSPSSGYMPMNOSSLGEACLDASVILGREGQFSRPLSL 1081  
QY 1104 HDPSLPQRYSEDTPLPSETDGVV-----APL-----TC-----SPOPE-----YVNOP 1143  
Db 1082 H-PIPRGR-----PASESEGHVTGSEAELEQKVCVRSRSPRPRGDSAYHSQR 1133  
QY 1144 DVROPSPREGP-----LPAARPAGATLERAKTLP--GKNGVW-----KDVLF 1185  
Db 1134 HSLLTPTVPLSPGPLEEDNGYVMPDTHLRGASSSREGTLLSVGLSSVLGTTEEDDED-- 1191  
QY 1186 AFGGAVENPEYLTPOGGAPQHP 1210  
Db 1192 -----EYEMNKRKRGSP-PRPP 1209

## RESULT 11

EGFR\_DROME STANDARD; PRT: 1426 AA.  
AC P04412; O61601; Q9W2G0; P81868;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)  
DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB).  
GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxId=7227;  
[1]  
RN  
RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
RX MEDLINE=94350209; PubMed=8070664;  
RA Clifford R., Schubach T.;  
RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
RT that several genetically defined classes of alleles cluster in  
RL subdomains of the receptor protein.";  
RL Genetics 137:531-550(1994).  
RN  
RP REVISIONS.

RA Clifford R., Schubach T.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
RX MEDLINE=85124611; PubMed=2982499;  
RA Livneh E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;  
RT "The Drosophila EGF receptor gene homolog: conservation of both  
RT hormone binding and kinase domains.";  
RL Cell 40:599-607(1985).  
RN [4]  
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
RX STRAIN=Oregon-R; TISSUE=Embryo;  
RX MEDLINE=87002474; PubMed=3093080;  
RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
RT "Alternative 5' exons and tissue-specific expression of the  
RT Drosophila EGF receptor homolog transcripts.";  
RL Cell 46:1091-1101(1986).  
RN [5]  
RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION ANALYSIS.  
RX MEDLINE=99102120; PubMed=9882502;  
RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;  
RT "Several levels of EGF receptor signaling during photoreceptor  
RT specification in wild-type, Ellipse, and null mutant Drosophila.";  
RL Dev. Biol. 205:129-144(1999).  
RN [6]  
RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
RX STRAIN=Berkely;  
RX MEDLINE=20196006; PubMed=10731132;  
RA Adams M.D., Calniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos B.L.G.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
RA Ballew K.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Kravitz S., Kulp D., Lai Z.,  
RA Foslter C., Gabriellian A.E., Garg N.S., Galbart W.M., Glasser K.,  
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulyov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.H.,  
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissenbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
RN [7]  
RP SEQUENCE OF 959-1078 FROM N.A.  
RX STRAIN=Daekwanryeong;  
RX MEDLINE=85137938; PubMed=2983232;  
RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;  
RT "A Drosophila genomic sequence with homology to human epidermal



RT growth factor receptor.";  
RL Nature 314:178-180(1985).  
RN [8]  
RP ANALYSIS.  
RX MEDLINE=92038942; PubMed=1936959;  
RA Raz E., Schejter E.D., Shilo B.Z.;  
RT "interallelic complementation among DER/fib alleles: implications for  
the mechanism of signal transduction by receptor-tyrosine kinases.";  
RL Genetics 129:191-201(1991).  
RN [9]  
RP REVIEW.  
RX MEDLINE=97248481; PubMed=9094709;  
RA Pertimon N., Perkins L.A.;  
RT "There must be 50 ways to rule the signal: the case of the Drosophila  
EGF receptor.";  
RL Cell 89:13-16(1997).  
CC -!- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
POLARITIES OF THE OOCYTE. IN THE EMBRIO, PLAYS A ROLE IN THE  
ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOEROSA  
AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
CUTICLE.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC -!- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND  
TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH  
IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
AND THORACIC AND ABDOMINAL GANGLIA.  
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF052754; AAC08536.1; -  
DR EMBL; AF052753; AAC08536.1; JOINED.  
DR EMBL; AF052754; AAC08535.1; -  
DR EMBL; AF052752; AAC08535.1; JOINED.  
DR EMBL; K03054; AAA51462.1; -  
DR EMBL; K03417; AAA51460.1; -  
DR EMBL; K03416; AAA50965.1; -  
DR EMBL; K03418; AAA51461.1; -  
DR EMBL; AF109077; AAD26134.1; -  
DR EMBL; AF109078; AAD26132.1; -  
DR EMBL; AF109082; AAD26132.1; JOINED.  
DR EMBL; AF109078; AAD26133.1; -  
DR EMBL; AF109084; AAD26133.1; JOINED.  
DR EMBL; AF109079; AAD26130.1; -  
DR EMBL; AF109081; AAD26130.1; JOINED.  
DR EMBL; AF109079; AAD26131.1; -  
DR EMBL; AF109083; AAD26131.1; JOINED.  
DR EMBL; AF109080; AAD26135.1; -  
DR EMBL; AE003454; AAF46732.1; -  
DR EMBL; X02293; CAA26157.1; -  
DR EMBL; X78920; CAA55523.1; -  
DR EMBL; X78918; CAA55521.1; -

DR EMBL; X78919; CAA55522.1; -  
DR PIR; A00640; GQFE.  
DR HSSP; P11362; LFCK.  
DR FlyBase; FBgn0003731; Egfr.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk pkkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr pkkinase.  
DR Pfam; PF00069; pkkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_pkkinase; 1.  
DR SMART; SM00261; FU; 7  
DR SMART; SM00219; TyrKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;  
KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;  
KW Developmental protein.  
FT SIGNAL 1 30  
FT CHAIN 31 1426  
FT DOMAIN 31 868  
FT TRANSMEM 869 889  
FT DOMAIN 890 1426  
FT DOMAIN 938 1198  
FT NP\_BIND 944 952  
FT BINDING 971 971  
FT ACT\_SITE 1063 1063  
FT MOD\_RES 902 902  
Query Match 28.6%; Score 1952; DB 1; Length 1426;  
Best Local Similarity 32.7%; Pred. No. 2.5e-96;  
Matches 468; Conservative 183; Mismatches 429; Indels 350; Gaps 42;  
QY 24 QVCTGDMKRLPASPEHDLMLRHLVQCQVQVQGNLELTLYPT-NASLSFLQDIOEVOG 82  
DB 100 KICIGTKSRLSPVSNKEHHYRLRDYNTCTVQGNLKLTLWLPNENLDSFLDNREVTG 159  
QY 83 YVLIHNOVQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTPTVTGASPGGL 137  
DB 160 YILISHVDVKKVFPKQLQIRGLTFLSLVSEEEKALFV-----TYSKM 203  
QY 138 RELQLRSITELKGVLLQRPOLQVQDTILWKDIFHKNNQLALTLDITNRSRACHPCSP 197  
DB 204 YLEIPDLRDVLNGQVGFHNNYNLCHMRTIQHSEIVSNGTDAYNYDFTAPECEPKCHE 263  
QY 198 MCKGSRGCGESDCCQSLTRTVCAGGCA--RCKGPLPTDCCHEQCAAGCTGPKHSDCLAC 255  
DB 264 SETHG-CWGEQPKNQKSKLTCSPQACAGRCYGPKECHLFCAGGCTGTQKDCIAC 322  
QY 256 LFNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTACPNYVLTSDVGSCTLVC 315  
DB 323 KNFFDEAVSKBCECPMRKYNPTTYVLETNPEKAYGATCVKECP-GHLLRDNGACVRSC 381  
QY 316 PLHNOEVTAEADGTQRCCKSKPCARVCYGLGMOYKANSKPIGITEL-----EPAGCKK 369  
DB 382 PQDKMDKGE-----CVPCNGPCPKTC-----PGVTVLHAGNIDSPENCTV 422  
QY 370 IFGSLAFLPESFDG--DPASNTA-----PLQEOLOVFEETLEETIGYLYISAWPDSLDP 421  
DB 423 IDGNIRILDQTFSGFDQDYANYTGMCPRIPLDPERREVFSTVKEITGYLVNIEGTHPOFRN 482  
QY 422 LSVFONLQVIRGRIHLNGAY-SLTLOGLIGISWLGRLSRLSGSLALIHNNHLCFVHTV 480  
DB 483 LSYFRNLETHGRQLMSFALAIVKSLVSLMRLNKQISSGSVWIGHNRDLVCVSN 542  
QY 481 PWDQLFRNPQALLHTANRPEDECVGEGLAHQLCARGHCWGPCTQCVNCSQFLRGQBC 540  
DB 543 RMPAIQKEPEQKVWVNNELRADLCEKNGTICSDQCNEBGCWAGDTQCLTCKNFNNGTC 602  
QY 541 VEECRVLQGLPREYVYNARHCLPCHPECPQNGSVTCFGEADQCVCACAHYKDPFCVARC 600

```
Db 603 IADCGYISNAYK--PDNRFTCKICHPECR-----TCNGAGADHCQECVHVRDQGHQCVSBC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNKYNDRGVRECHATCGDCTGPKDTIGIGACTTCNLA1INNDATVVKRCLLKDDKCPD- 713
Qy 609 SYMPTWKF--PDEEGACQP-----CPI-----NCTH-----632
Db 714 GY--FMVEYVHPOEGSLKPLAGRAVCRKCHPLCELCTNYGYHEQVCSKCTHYKRRQCE 771
Qy 633 -----SC-----VDLDKDG-----641
Db 772 ECPADHYTDEORECFORHEPCNGCTGPGADCKSCRNFKLFDANETGYVNSTWPNCT 831
Qy 642 -CPAQR-----ASPLTS-----IVSAVVGILLVVVLGVVFGI 673
Db 832 KCPLEMRHNYOVYTAIGPYCAASPPRSSKITANLDVNMFIITGAVLVPTICILCV--T 889
Qy 674 LIKRQOKIRKYT--MRRLLOETELVEPLTPSGMNPQAOIRLAKETELRKVKVLGSGAF 731
Db 890 YICROKQAKKETVKNMTALSCEDSEPLRPSNIGANLCKLRIVDAELRKGVLGMGAF 949
Qy 732 GTVYAGIWIPIGDNENKVIPIVAIKVLRENTSPKANKEILDEAYVMAGVGSPPVGRLLIGICLT 791
Db 950 GRVYGVVPEGENKVIPIVAIKELLKSTCAESSEEFLEAYIMASEHVNLLKLLAVCMS 1009
Qy 792 STVQLVTQMPYGCCLLDHVRNRLGSLDNLNMCQIAKMSYLEVDVRLVHRDLAARNV 851
Db 1010 SQMMLITQMPGLCLLDVYVNRNRDKIGSKALLNWSQIAKMSYLEEKLVRDLAARNV 1069
Qy 852 LVKSPNHVKITDFGLARLLDIDETEHADGGKVPKWMMALESILRRRFTHQSDVWSYGV 911
Db 1070 LVQTPSLVKITDFGLAKLLSDSNEYKAAAGGRKPIKWLALECIRNRVFTSKSDVWAF 1129
Qy 912 VNELMTFGAKPYDGIPIAREIPDLLEKGERLPPICTIDVYIMVWKWMIDSECRPRE 971
Db 1130 IWELTTFGQPHENIPAKDIPDLIEVLGKLEQPEICSLDIYCTLLSCHWHLDAAMRTFKQ 1189
Qy 972 LVSEFSRMARDPQRFVVIQNEIDL--PASPLDSTFYRSILLEDD--DMGDLDVAEYLVP 1026
Db 1190 LTTVFAEFAFGVLAIPGDKFTRLPA-----YTSQDEKDLIRKLAPTPTDGEAIAK 1242
Qy 1027 QGFFCPDPAPGAGGVHHRSSSTRSGGDLTLGLEPSEAP-----RSLPAPSEG 1080
Db 1243 PDYIOPKAAAGPS-----HRTDCT-----DEMPKLNRYCKDPSKNKSS 1281
Qy 1081 AGSDVFDG---DLGMAAGKGLSLPHTDPSPLQRYSEDPTVPLPSETDGYVAPLTCSPOP 1137
Db 1282 TGDERSDSSAREVGGNLR-----LDLPVEDDDYLMPTCQOP 1319
Qy 1138 EYVNPQDVRPQPSREGPLPAARPAATLERAKTLPQKNGVVKDVFAFGAVENPEYL 1197
Db 1320 NNNNNNN-----NPNQNNMAAVGAAGYM-----DLIGVPVSDNPEYL 1358
Qy 1198 ----TPQGAAPQPH-----PPAFSP-AFDNLYYWD 1224
Db 1359 LNAQTLGVGESPIQTIGIPVWGGETVMEVKVPMPSSEPTSSDHEYND 1408
```

## RESULT 12

```
ERBB_ALV STANDARD; PRT; 634 AA.
ID ERBB_ALV
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nilsson T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726 (1985).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -!- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M10066; AAA48763.1; ALT_INIT.
DR PIR; A00643; TVCHLV.
DR PIR; B00643; TVFLV.
DR HSP; P11362; IFKG.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.
DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
```

Query Match 25.7%; Score 1749.5; DB 1; Length 634;  
Best Local Similarity 52.3%; Pred. No. 5.7e-86;  
Matches 370; Conservative 79; Mismatches 135; Indels 123; Gaps 17;

```
Qy 587 CAHYKDPFPCVACPSGVKPDLSYMPIMKFPDEEGACQPCINCTHSCVDLDDKGCPEAQ 646
Db 3 CAHFDIGHCVKACFAGVLGENDTL-VMKYADANAVCQLCHPNCTRGCKPGLEGCP--- 58
Qy 647 RASPLTSTIVSAVV-GILLVVVLGVVFGIILKROOKIRKYTMRLRLQETELVEPLTPSGA 705
Db 59 NGSKTPSAAGVVGILLVVLGVVGLGILVLR--HIVKRTLRRLQERELVEPLTPSGE 117
Qy 706 MPNQAMRLKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANK 765
Db 118 APNQAHRLKETEFKVKVGLGSGAFGVYKGLWIPEGEKVKIPVAIKELREATSPKANK 177
Qy 766 EILDAYVMAGVSPVSRLLGICLTSTVLTQMPYGCCLLDHVRNRLGSLDNLNMCQIAKMSYLEVDVRLVHRDLAARNV 825
Db 178 EILDAYVMASVDNPHVCRLLGICLTSTVLTQMPYGCCLLDYREHKDNTGYSQYLLNW 237
Qy 826 CNQIAKMSYLEVDVRLVHRDLAARNVVKSPNHVKITDFGLARLLDIDETEHADGGKVP 885
Db 238 CVQIAKGNVLEERLVRDLAARNVVKTPQHVKITDFGLAKLLGADKEVHAEKGVKVP 297
Qy 886 IKWMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 945
```

```
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYMTVMKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMTVMKCMWIDADSRPKFRELIAEFSKWARDPPRYLVIOQDERMHLSPDTSKF 417
Qy 1005 YRSLLDDMDGLVDAAEYLVPOQGFCDPAPGAGGMVHRRHSSTRSGGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRPL-----APSEGAGSDVFDGLGMAAKGLQSLTPHDPSPLOQRYSEPTVP 1119
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHFVREDTSFQRYSSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQEVYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTILSPGK 1177
Db 496 FLEESIDDGFL-----PAPEYVNO--LMPKKFS-----TAMVQ 526
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPAFSPAFDNLY 1222
Db 527 NQIYNNISUTAIKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSPY 578
Qy 1223 WDQ-----DPPE-----RCAPPSTFKGTPTAENPEYLGIDVP 1254
Db 579 WTQSGNHQINLNDPDYQODFLPNETKPNGLLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
CT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 01, Last annotation update)
DE Tyrosine-protein kinase transforming protein erBB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain E54).
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erBB gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=84223957; PubMed=6328659;
RA Debuire B., Henry C., Benaissa M., Biserte G., Claverie J.-M.,
RA Saule S., Martin P., Stehelin D.;
RT "Sequencing the erba gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -!- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
DR EMBL; K02006; AAA42394.1; ALT_INIT.
DR EMBL; K01216; AAA42400.1; -.
DR PIR; A00644; TVYUH.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00219; TyrKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399 PROTEIN KINASE.
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MH; 768BCDD06745D609 CRC64;

Query Match 25.0%; Score 1703; DB 1; Length 604;
Best Local Similarity 52.2%; Pred. No. 1.6e-83;
Matches 360; Conservative 76; Mismatches 128; Indels 126; Gaps 16;

Qy 587 CAHYKDPPEVCARCPGKPDLSYMPINKFPDEEGACQPCINCHTSCVDLDDKGCPAEQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL--VRKYADANAVCOLCHPNCTRGCKPGLEGCCP--- 58
Qy 647 RASPLTISVAVV--GILLVVLGVVFGILIKRRQKIRKYMRRLQETELVEPLTPSGA 705
Db 59 NSKTPSIAAGVVGGLLCLVVGGLGILVLR--HIVKRTLRRLQERLEPLTPSGE 117
Qy 706 MPNQAQMRILKETELRKVKVLGSGAGFTYVKGHWIPDGENVKIPVAIKVLENTSPKANK 765
Db 118 APNQAHLRIKETEFKKVKVLGSGAGFTYVKGHWIPDGENVKIPVAIKVLENTSPKANK 177
Qy 766 EILDEAYVMAGVSPVSRLLGCLTSTVOLTPMLPYGCLLDHVRNCRGLCSODLLNW 825
Db 178 EILDEAYVMASVDNPHVCRLGLCLTSTVOLTPMLPYGCLLDYIREHKDNIGSQYLLNW 237
Qy 826 CMQIAGKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVP 885
Db 238 CVQIAGKMYLERRLVHRDLAARNVLVKTPOHKITDFGLAKLLGADEKEYHAEGKVP 297
Qy 886 IKWMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQP 945
Db 298 IKWMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIPIASEISSVLEKGERLPQP 357
Qy 946 ICTIDVYMTVMKCMWIDSECRPRFRELVSFSESMARDPQRFVVIQ-NEDLGPASPLDSTF 1004
Db 358 ICTIDVYMTVMKCMWIDADSRPKFRELIAEFSKWARDPPRYLVIOQDERMHLSPDTSKF 417
Qy 1005 YRSLLDDMDGLVDAAEYLVPOQGFCDPAPGAGGMVHRRHSSTRSGGGDLTLGLE 1064
Db 418 YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST----- 449
Qy 1065 PSEEEAPRPL-----APSEGAGSDVFDGLGMAAKGLQSLTPHDPSPLOQRYSEPTVP 1119
Db 450 -----SRTPLLSLSLTSATNSATNCID-----RNGQGHFVREDTSFQRYSSDPTGN 495
Qy 1120 LPSET--DGYVAPLTCSPQEVYVQPDVVRPQPPSPREGPLPAARPAGATLERAKTILSPGK 1177
Db 496 FLEESIDDGFL-----PAPEYVNO--LMPKKPSTAM----- 524
Qy 1178 NGVVKDVF-----AFGGAIVENPEYLTPOGGAAPQPHPPAFSPAFDNLY 1222
Db 525 --VQNIYNNISUTAIKLPMDSRYSQNSHSTAVDNPEYL-----NTNQSPLAKTVFPE 574
Qy 1219 NLIYVWDQDPPERGAPPSTFKGTPTAENPEY 1248
```

Db	178	EILDEAYVASVDNPHVCRLLGLICLTSTVQLITQILMPYGCGLLDYIREHKDNICSGYLLNW	237
Qy	826	CMQIAKMGSYLEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVP	885
Db	238	CVQIAKMGNYLEBRHVMVHRDLAARNLVKTPQHVKITDFGLAKQLGADEKEYHAEGGKVP	297
Qy	886	IKWMALESILRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPOPP	945
Db	298	IKWMALESILHRTYTHQSDVWSYGVTVWELMTFGSKPYDGIPIASEISVLEKGERLPOPP	357
Qy	946	ICTIDVYMLVWKCMWIDSECRPRFRELVSFEFSMARDPQRFVVIQ-NEDLGPASPLDSTF	1004
Db	358	ICTIDVYMLVWKCMWDASGRPKFERELIAEFESKWARDPPRYLVIQGDERVHLPSPTDSKF	417
Qy	1005	YRSLLEDDDDGDLVDAAEYLVPOQGFCDPAPGAGGMVHHRHSSTRSGGGDLTLGLE	1064
Db	418	YRTLMEEDMEDIVDAEYLVPHQGF-----NSPST-----	449
Qy	1065	PSEEBAPRSP-----APSEGAGSDVFDGDLGCAAKQLSLTHDPSPLQRYSEDTVP	1119
Db	450	-----SRTPLLSSLSATSNNSATNCIDRNG-----H-----	476
Qy	1120	LPSETGYVAPLTCSPQPEYVNOVDVVRPPSPREGPLPAARPAGAT-LERAKTLPSPGN	1178
Db	477	-PVREDGFL-----PAPEYVNO--LMPKKPSTAMVQNIYVYISLTATSKLPIDSKRYQN	527
Qy	1179	GVVKDVFAFGGAVENPEYL 1197	
Db	528	-----SHSTAVDNP EYL 539	
RESULT 15			
EGFR_CHICK			
ID	EGFR_CHICK	STANDARD;	PRT; 703 AA.
AC	P13387;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JAN-1990 (Rel. 13, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER		
DE	(Fragment).		
GN	EGFR.		
OS	Gallus gallus (Chicken).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Aukosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;		
OC	Gallus.		
OX	NCBI_TaxID=9031;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88261272; PubMed=3260329;		
RA	Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,		
RA	Ullrich A., Vennstrom B., Schlesinger J., Givol D.;		
RT	"Chicken epidermal growth factor (EGF) receptor: cDNA cloning,		
RT	expression in mouse cells, and differential binding of EGF and		
RT	transforming growth factor alpha."		
RL	Mol. Cell. Biol. 8:1970-1978(1988).		
CC	-1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,		
CC	AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND		
CC	VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).		
CC	-1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-1- SUBCELLULAR LOCATION: Type I membrane protein.		
CC	-1- MITOCELLANEOUS: Binding of EGF to the receptor leads to		
CC	dimerization, internalization of the EGF-receptor complex,		
CC	induction of the tyrosine kinase activity, stimulation of cell DNA		
CC	synthesis, and cell proliferation.		
CC	-1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announcement/">http://www.isb-sib.ch/announcement/</a>		

or send an email to license@sib.ch).

CC EMBL; M20386; AAA48760.1; ..  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR In-erPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR SMART; SM00261; FU; 4.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; PARTIAL.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; PARTIAL.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; PARTIAL.  
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
FT SIGNAL 1 30  
FT CHAIN 31 >703 EPIDERMAL GROWTH FACTOR RECEPTOR.  
FT DOMAIN 31 654 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 655 667 POTENTIAL.  
FT DOMAIN 668 >703 CYTOPLASMIC (POTENTIAL).  
FT DISULFID 197 206 BY SIMILARITY.  
FT DISULFID 201 214 BY SIMILARITY.  
FT DISULFID 222 230 BY SIMILARITY.  
FT DISULFID 226 238 BY SIMILARITY.  
FT DISULFID 239 247 BY SIMILARITY.  
FT DISULFID 243 255 BY SIMILARITY.  
FT DISULFID 258 267 BY SIMILARITY.  
FT DISULFID 271 298 BY SIMILARITY.  
FT DISULFID 302 314 BY SIMILARITY.  
FT DISULFID 318 333 BY SIMILARITY.  
FT DISULFID 336 340 BY SIMILARITY.  
FT DISULFID 513 522 BY SIMILARITY.  
FT DISULFID 517 530 BY SIMILARITY.  
FT DISULFID 533 542 BY SIMILARITY.  
FT DISULFID 546 562 BY SIMILARITY.  
FT DISULFID 565 581 BY SIMILARITY.  
FT DISULFID 569 589 BY SIMILARITY.  
FT DISULFID 592 601 BY SIMILARITY.  
FT DISULFID 605 627 BY SIMILARITY.  
FT DISULFID 630 638 BY SIMILARITY.  
FT DISULFID 634 646 BY SIMILARITY.  
FT CARBOHYD 134 134 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 190 190 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 359 359 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 368 368 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 420 420 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 573 573 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 578 578 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 613 613 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 633 633 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 648 648 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT NON\_TER 703 703  
SQ SEQUENCE 703 AA; AFF2DE11B735A690 CRC64;

Query Match 23.4%; Score 1595; DB 1; Length 703;

Best Local Similarity 44.6%; Pred.No.le-77;  
Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;

QY 8 RAGLLALLPPGAA-----STQVCCTGTDMLKLPASPETHDMLRHLYQGCVOVQGNLE 61  
Db 13 RGAUVLLVLLGVALCSAVEEKKVCOGTNNKLTQLGHVEDFTSLQRMYNCEWLSNLE 72  
QY 62 LTYLPTNASLSLQDIQEVQGVYLIANQVQVLPQRLRIVRGTLQFDNYALAVLDNGD 121  
Db 73 ITYVEHNRDLTLKTIQEVAGYVLIANVVDVIPLENLQIIRGNVLYDNSFALAVLSNTH 132  
QY 122 PLNNTTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPOLCYODTILWKDIIFHKNNQLAL 181  
Db 133 -NNKTO-----GLRELPMKRLSELNGVVKISNNPKLCNNDTVLWNDIIDSRSK-PL 182  
QY 182 TLID-TNRSRACHPCSPCKSGRCWSESSEDQSLTRTVACGCA-RCKGPLPTDCCHEQ 239  
QY 182 TLID-TNRSRACHPCSPCKSGRCWSESSEDQSLTRTVACGCA-RCKGPLPTDCCHEQ 239

Db 183 TVLDFASNLSSCPKCHPNCTEDHMGAGEBQNCOTLTKVICAQCSGRCGRKVPSPDCCHNQ 242  
QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTPESNPNPEGRTVTCASCVTAC 299  
Db 243 CAAGCTGPPRESCLACRKRFRDATTCTPPLVLYNPTTYQMDVNPBEGKYSFGATCVREC 302  
QY 300 PYNILSTDVGSCTLVCPHNEVTAEDGTQRCCEKSKPCARVCYGLGMOYIKANSKFIGI 359  
Db 303 PHNYVVDHGSVCVRSCTNTDYEY-EENGVRKCKKCDGLCKVCNCGIGIGELKGILS-INA 360  
QY 360 TELE-PAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLBEITGYLISAMPDS 418  
Db 361 TNIDSPKCTKINGDVSIPLVAFGLDAFTKTLPLDPKLDVFRVKEISGFLLIQAMPDN 420  
QY 419 LPDLSVFONLQVIRGRIHNGAYSLTLOGLGISWGLRSILRELGSGLALIHNNHLCFVH 478  
Db 421 ATDLVAFENLEIIRGRTKQHQYSLAVNKLQSLGLRLSKEISDGDIAIMKNKLCYAD 480  
QY 479 TVPWDLFRPHQALLHTANRPEDECVGEGELACHQLCARGHGWPGPTQCVNCSQFLRGQ 538  
Db 481 TNWRSLFATQSKTKIIQNRNKNDCDTRHVCDDPLCSDVGCWGPFPFHCFSRFFSRQK 540  
QY 539 ECVEECRVLQGLPREYVNAHCLPCHPECOQNG----SVTCFQPEADQCACAHYKDPDF 595  
Db 541 ECVKQCNILQGBPREPERDSKCLPCHSECLVONSTAYNTTCSGPGPDHCKCAHFDGPH 600  
QY 596 CVARCPGKPDLSYMPIWKFPDEGACQPCPINCTHSCVDLDDKCPAEQORASPLTSIV 655  
Db 601 CVKACPAVGLGENDTL-VMKYADANAVCOLCHPNCTRCKGKPGLEGCP---NGSKTPSIA 656  
QY 656 SAVV-GILLVVVGLVVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTP 702  
Db 657 AGVVGSLCLVVVGLGILGLYLRRL-HIVKRTLLRLLQERELVEPLTP 703

Search completed: July 22, 2003, 08:45:48

Job time : 20.2304 secs

GenCore version 5.1.6  
Copyright (C) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds  
(without alignments)  
4403.399 Million cell updates/sec

Title: SEQ4-103-117-14

Perfect score: 6847

Sequence: 1 MELAALCRWGLLLALLPGCA.....TFKGTPTAENPEYLGIDVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

```

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6714	98.1	1255	21	Human heregulin 2
2	6714	98.1	1255	22	AAE12130
3	6714	98.1	1255	22	HER2 transgene kin
4	6714	98.1	1255	23	AAU74545
5	6708	98.0	1255	17	AAW01111
6	6708	98.0	1255	20	AAW92406
7	6708	98.0	1255	21	AAW21198
8	6708	98.0	1255	21	AAW84780
9	6708	98.0	1255	22	AAW85458
10	6708	98.0	1255	22	AAW88267

11	6708	98.0	1255	23	AAE24067	Human Her-2 protei
12	6708	98.0	1255	23	AAE20479	Human Her-2/neu pr
13	6708	98.0	1255	23	AAW51143	Human Her-2/neu on
14	6708	98.0	1255	23	AAU77114	Human Her-2/neu po
15	6665	97.3	1433	14	AAK39568	Sequence of c-erbB
16	6544	95.6	1223	13	AAU98923	Human breast cance
17	6391	93.3	1200	21	AAW21208	Human HER-2/neu pr
18	5950.5	86.9	1256	21	AAW21199	Rat Her-2/neu prot
19	5950.5	86.9	1256	23	AAW51144	Rat Her-2/neu onco
20	5918.5	86.4	1256	21	AAW21206	Mouse Her-2/neu pr
21	5918.5	86.4	1256	22	AAW62860	Amino acid sequenc
22	5918.5	86.4	1256	23	AAW51151	Human HER-2/neu on
23	4794	70.0	919	21	AAW21203	Mouse Her-2/neu fu
24	4794	70.0	919	23	AAW51148	Her-2/neu extracel
25	4068.5	59.4	920	23	AAW51152	Mouse Her-2/neu ex
26	4068.5	59.4	926	23	AAW51153	Mouse Her-2/neu ex
27	3678	53.7	712	21	AAW21204	Human HER-2/neu fu
28	3678	53.7	712	23	AAW51149	Her-2/neu extracel
29	3532	51.6	782	18	AAW19764	Her2-GM-CSF immuno
30	3530	51.6	653	21	AAW21200	Extracellular HER-
31	3530	51.6	653	23	AAW51145	Human Her-2/neu on
32	3492	51.0	645	22	AAW60408	Human ErbB2 oncopr
33	3492	51.0	645	22	AAW61593	Human ErbB2 extrac
34	3427	50.1	951	21	AAW44993	DC9cFv-erbB2EC fu
35	3326	48.6	624	11	AAW08222	Extracellular port
36	3139	45.8	1210	21	AAW19259	Amino acid sequenc
37	3139	45.8	1210	21	AAW50616	Human EGF receptor
38	3139	45.8	1210	23	AAE23019	Human Her-1 protei
39	3139	45.8	1210	23	AAW50768	Human epidermal gr
40	3137	45.8	1210	22	AAW68420	Amino acid sequenc
41	3098	45.2	1210	22	ABP51768	Human epidermal gr
42	3084	45.0	583	23	AAE20483	Human protein for
43	3084	45.0	587	23	AAE20481	Human protein for
44	3083	45.0	589	23	AAE20484	Human protein for
45	3083	45.0	600	23	AAE20482	Human protein for

#### ALIGNMENTS

RESULT 1  
AAW92620  
ID AAW92620 standard; Protein; 1255 AA.  
XX  
AC AAW92620;  
XX  
DT 10-AUG-2000 (first entry)  
XX  
DE Human heregulin 2 (Her2).  
DE  
KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;  
KW self-protein; cancer; breast cancer; prostate cancer;  
KW cell-associated peptide antigen; foreign epitope.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT Domain  
FT Location/Qualifiers  
FT /label= N-terminal  
FT /note= "mature polypeptide"  
FT 5..25  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 59..73  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 103..117  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 149..163  
FT /label= insertion region  
FT /note= "suitable for foreign epitope insertion"  
FT 174..323



QY 777 AYVMAGVSPYVSRLLGICLTSTVOLVTPQMPYGCLLDHVRENRRGLSQDILLNWCMTA 836  
 DB |||||  
 771 AYVMAGVSPYVSRLLGICLTSTVOLVTPQMPYGCLLDHVRENRRGLSQDILLNWCMTA 830  
 DB |||||  
 QY 837 KGMYSLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIETEHADGKVPKWA 896  
 DB |||||  
 831 KGMYSLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDDIETEHADGKVPKWA 890  
 DB |||||  
 QY 897 LESILRRRFTHOSDWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTID 956  
 DB |||||  
 891 LESILRRRFTHOSDWSYGVTVWELMTGAKPYDGI PAREIPDLLEKGERLPQPICTID 950  
 DB |||||  
 QY 957 VYIMVKKWMDISECRPRFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1016  
 DB |||||  
 951 VYIMVKKWMDISECRPRFRELVSFSSRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1010  
 DB |||||  
 QY 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEA 1076  
 DB |||||  
 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRHRSSTRSGGDLTLGLEPSEEA 1070  
 DB |||||  
 QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLTPHPSPLQRYSEPTVPLPSETDGYVAP 1136  
 DB |||||  
 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLTPHPSPLQRYSEPTVPLPSETDGYVAP 1130  
 DB |||||  
 QY 1137 LTCSPQPEVYNQDVRPQPPSPREGPLPAARAGATLERAKTLPSPKNGWKDVAFGGA 1196  
 DB |||||  
 1131 LTCSPQPEVYNQDVRPQPPSPREGPLPAARAGATLERAKTLPSPKNGWKDVAFGGA 1190  
 DB |||||  
 QY 1197 VENPEYLTPOGGAAPQPHPPAFSPAFNLYYWDQDPPERGAPPTFKGTPTAENPEYLG 1256  
 DB |||||  
 1191 VENPEYLTPOGGAAPQPHPPAFSPAFNLYYWDQDPPERGAPPTFKGTPTAENPEYLG 1250  
 QY 1257 LDVPV 1261  
 DB |||||  
 1251 LDVPV 1255  
 DB |||||

RESULT 2  
 AAEL12130  
 ID AAEL12130 standard; Protein; 1255 AA.  
 XX  
 AC AAEL12130;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human tyrosine kinase-type receptor, HER-2.  
 XX  
 KW Therapeutic compound; major histocompatibility complex; vaccine;  
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;  
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
 KW antigen presenting cell; human; tyrosine kinase-type receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 Key Location/Qualifiers  
 FT 774...782  
 FT /note= "Antigenic epitope"  
 XX  
 FN WO200168677-A2.  
 XX  
 PD 20-SEP-2001.  
 XX  
 PF 16-MAR-2001; 2001WO-US40328.  
 XX  
 PR 16-MAR-2000; 2000US-0527487.  
 XX  
 PA (GENZ ) GENZYME CORP.  
 XX  
 PI Nicolette CA;  
 XX  
 DR WPI; 2001-616284/71.  
 DR N-PSDB; AAD19731.  
 XX

PT Novel synthetic therapeutic compound for inducing immune response and  
 PT for use in adoptive immunotherapy, has enhanced binding to major  
 PT histocompatibility molecules and enhanced immunoregulatory properties  
 PT  
 PS  
 PS  
 XX Claim 4; Page 63-67; 69pp; English.  
 XX  
 CC The invention relates to synthetic therapeutic compounds (antigenic  
 CC peptides) with enhanced binding to major histocompatibility complex  
 CC (MHC) molecules and enhanced immunoregulatory properties relative  
 CC to their natural counterparts. Compounds of the invention are useful  
 CC for inducing an immune response in a subject and for use in adoptive  
 CC immunotherapy. They are useful as components of anti-cancer vaccines  
 CC and to expand immune effector cells that are specific for cancers  
 CC characterised by expression of the breast cancer antigen, HER-2.  
 CC Polynucleotides that encode peptides of the invention are useful as  
 CC hybridisation probes and as primers for the detection of genes of gene  
 CC transcripts that are expressed in antigen presenting cells (APCs), to  
 CC confirm transduction of polynucleotides into host cells. The present  
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
 CC of the invention are designed based on the HER-2 antigenic peptide  
 CC (774-782).  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 98.1%; Score 6714; DB 22; Length 1255;  
 Best Local Similarity 98.3%; Pred. No. 0;  
 Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;  
 QY 1 MELAALCERWGLLLALLPGAASTQVCTGTDKMLRLPASPETHLDMRLHYQSCVVOGNL 60  
 DB |||||  
 1 MELAALCERWGLLLALLPGAASTQVCTGTDKMLRLPASPETHLDMRLHYQSCVVOGNL 60  
 DB |||||  
 QY 61 ELTYLPTNASLSFLODIOEVQGYVLIANQVQVPLQRLRIVRGTQLPEDNVALAVLDNG 120  
 DB |||||  
 61 ELTYLPTNASLSFLODIOEVQGYVLIANQVQVPLQRLRIVRGTQLPEDNVALAVLDNG 120  
 DB |||||  
 QY 121 DPLNNFNFTVFWLRVPKVSAS-----HLEQLRSLEILKGGVLIQRIQNPQCYQDTILWK 176  
 DB |||||  
 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQRIQNPQCYQDTILWK 170  
 DB |||||  
 QY 177 DIFHKNQALTLIDTNRSRACHPCSPKSGRCHGESSEDCOSLTRTVCAGGCARCKGP 236  
 DB |||||  
 171 DIFHKNQALTLIDTNRSRACHPCSPKSGRCHGESSEDCOSLTRTVCAGGCARCKGP 230  
 DB |||||  
 QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRYT 296  
 DB |||||  
 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRYT 290  
 DB |||||  
 QY 297 FGASCVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLMEHL 356  
 DB |||||  
 291 FGASCVTACPNYLSLTDVGSCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLMEHL 350  
 DB |||||  
 QY 357 REVRAVTSANTQEFAGCKKIFGSLAFLPESDGDPAASNTAPLQPOLQVFTLEETIGYL 416  
 DB |||||  
 351 REVRAVTSANTQEFAGCKKIFGSLAFLPESDGDPAASNTAPLQPOLQVFTLEETIGYL 410  
 DB |||||  
 QY 417 YISAWPDSLPLDSVFNQVIRGRILHNGAYSLTIQGLIGISWLGRLSRELGSGLALIH 476  
 DB |||||  
 411 YISAWPDSLPLDSVFNQVIRGRILHNGAYSLTIQGLIGISWLGRLSRELGSGLALIH 470  
 DB |||||  
 QY 477 NTHLCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHMGPGTQCVN 536  
 DB |||||  
 471 NTHLCFVHTVPDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGHMGPGTQCVN 530  
 DB |||||  
 QY 537 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECOPOGSGVTCFGEADQCACAHY 596  
 DB |||||  
 531 CSQFLRGQECVEECRVLOGLPREYVNAHCLPCHPECOPOGSGVTCFGEADQCACAHY 590  
 DB |||||  
 QY 597 KDPFFCVARCPGSKVPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGPAPFORASP 656  
 DB |||||  
 591 KDPFFCVARCPGSKVPDLSYMPIWKFPDEEGACOPCPINCTHSCVDLDDKGPAPFORASP 650  
 DB |||||



```
QY 657 LTSIVSAVVGILLVVVGVVFGILIKRROOKIRKVTMRRLLOETELVELPTSGAMPNQA 716
DB 651 LTSIVSAVVGILLVVVGVVFGILIKRROOKIRKVTMRRLLOETELVELPTSGAMPNQA 710
QY 717 QMRILKETELRKVKVLSGAGFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 776
DB 711 QMRILKETELRKVKVLSGAGFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 770
QY 777 AYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHHVNRGRGLSGODLLNMCQIA 836
DB 771 AYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHHVNRGRGLSGODLLNMCQIA 830
QY 837 KCMYSLEDVRLVHRDLAARNVLKSPNHNKIIDFGLARLLDIDETEHADGKGKVPWKMA 896
DB 831 KCMYSLEDVRLVHRDLAARNVLKSPNHNKIIDFGLARLLDIDETEHADGKGKVPWKMA 890
QY 897 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
DB 891 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
QY 957 VYMIWVKWMIIDSECRPRELVSFSEFRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
DB 951 VYMIWVKWMIIDSECRPRELVSFSEFRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
QY 1017 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGGVHHRHRSSTRSGGGDLTLGLPSEEEA 1076
DB 1011 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGGVHHRHRSSTRSGGGDLTLGLPSEEEA 1070
QY 1077 PRSPLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAP 1136
DB 1071 PRSPLAPSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPCKNGVWVDVAFGGA 1196
DB 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPCKNGVWVDVAFGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLYWDQDPPPERGAPPSTFKGTPTAENPEYLG 1256
DB 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLYWDQDPPPERGAPPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 3
AAB60167
ID AAB60167 standard; Protein; 1255 AA.
AC AAB60167;
DT 03-APR-2001 (first entry)
DE HER2 transgene plasmid construct encoded protein.
KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;
KW antibody.
XX Homo sapiens.
OS Synthetic.
PN WO200100244-A2.
PD 04-JAN-2001.
XX 23-JUN-2000; 2000WO-US17229.
PR 25-JUN-1999; 99US-0141316.
PR 16-MAR-2000; 2000US-0189844.
XX (GenT) GENENTECH INC.
XX Erickson S, Schwall R;
PI
```

```
XX WPI: 2001-061962/07.
DR N-PSDB; AAF24297.
XX
PT Treating tumors, particularly breast cancers, which overexpress an ErbB
PT receptor and does not respond to an anti-ErbB antibody, comprises
PT conjugating the antibody to a maytansinoid -
XX
PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 98.1%; Score 6714; DB 22; Length 1255;
Best Local Similarity 98.3%; Pred No. 0;
Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;
QY 1 MELAALCRWGLLLALLPFGAASQVCTGTDMLRLPASPEHLDMLRLHYQGCQVVGNNL 60
DB 1 MELAALCRWGLLLALLPFGAASQVCTGTDMLRLPASPEHLDMLRLHYQGCQVVGNNL 60
QY 61 EUTYLPNTASLSFLODIOEVQGYVLIHNVQVQVPLQRLIRVGTQOLFEDNALAVLDNG 120
DB 61 EUTYLPNTASLSFLODIOEVQGYVLIHNVQVQVPLQRLIRVGTQOLFEDNALAVLDNG 120
QY 121 DPLNNFNFTVSFMLRVPKVSAS----HLLEQLRSLEILKGGVLIORNPOLCYODTILWK 176
DB 121 DPLNN-----TTPVTGASPGGLRELQRLSLEILKGGVLIORNPOLCYODTILWK 170
QY 177 DIFHKNNQALALTLIDNRSRACHPCSPCKSGRCWGESSEDCOSLTRTVCAAGCARCKGP 236
DB 171 DIFHKNNQALALTLIDNRSRACHPCSPCKSGRCWGESSEDCOSLTRTVCAAGCARCKGP 230
QY 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
DB 231 LPTDCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290
QY 297 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEQGTORCEKCKPCARVCYGLGMEHL 356
DB 291 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEQGTORCEKCKPCARVCYGLGMEHL 350
QY 357 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPEQLQVFETLEEITGYL 416
DB 351 REVRAVTSANIOEFAGCKKIFGSLAPLPSFDGDPASNTAPLOPEQLQVFETLEEITGYL 410
QY 417 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLSGLALIIH 476
DB 411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSRLSGLALIIH 470
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPDECVCGEGLACHOLCARGHCWGPGTQCVN 536
DB 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPDECVCGEGLACHOLCARGHCWGPGTQCVN 530
QY 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHY 596
DB 531 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADOCVACAHY 590
QY 597 KDPFFCVARCPGKVPDLSYMPIWKFPPDEBEGACQPCPINCTHSCVLDLDDKGCPAEORASP 656
DB 591 KDPFFCVARCPGKVPDLSYMPIWKFPPDEBEGACQPCPINCTHSCVLDLDDKGCPAEORASP 650
QY 657 LTSIVSAVVGILLVVVGVVFGILIKRROOKIRKVTMRRLLOETELVELPTSGAMPNQA 716
DB 651 LTSIVSAVVGILLVVVGVVFGILIKRROOKIRKVTMRRLLOETELVELPTSGAMPNQA 710
QY 717 QMRILKETELRKVKVLSGAGFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 776
DB 717 QMRILKETELRKVKVLSGAGFCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 776
```

```
Db 711 QNRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKILDE 770
Qy 777 AYVMAGVSPYVSRLGIGICTSTVQLVTLQMPYGCCLLDHVRENRRGLSGQDLNLCWQIA 836
Db 771 AYVMAGVSPYVSRLGIGICTSTVQLVTLQMPYGCCLLDHVRENRRGLSGQDLNLCWQIA 830
Qy 837 KGSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKQMA 896
Db 831 KGSYLEDLVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGGKVPKQMA 890
Qy 897 LESILRRRTHOSDVMWSYGVYWMELTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRTHOSDVMWSYGVYWMELTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
Qy 957 VYMIWVKWMIIDSECRPRRELVSERPMARDPQRFVWIONEDLGPASPLDSTFYRSLLLE 1016
Db 951 VYMIWVKWMIIDSECRPRRELVSERPMARDPQRFVWIONEDLGPASPLDSTFYRSLLLE 1010
Qy 1017 DDDMGDLVDAEYLVLPQQGFCCPDPAAGGMMVHRRSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVLPQQGFCCPDPAAGGMMVHRRSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSGAGSDVDGDLGMAAGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAP 1136
Db 1071 PRSPLAPSGAGSDVDGDLGMAAGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAP 1130
Qy 1137 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAATLERAKTSLSPGKNGVWVDVAFGGA 1196
Db 1131 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAATLERAKTSLSPGKNGVWVDVAFGGA 1190
Qy 1197 VENPEYLTQGGAAPOPHPPAPSPAFDNLYWDQPPRGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTQGGAAPOPHPPAPSPAFDNLYWDQPPRGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 4
AAU74545
ID AAU74545 standard; Protein; 1255 AA.
XX
AC AAU74545;
XX
DT 23-APR-2002 (first entry)
XX
DE Human HER2 (ErbB2) polypeptide.
XX
KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;
KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;
KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;
KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;
KW glial disorder; astrocytal disorder; hypothalamic disorder;
KW glandular disorder; macrophagal disorder; epithelial disorder;
KW stromal disorder; blastocoealic disorder; inflammatory disorder;
KW angiogenic disorder; immunological disorder.
XX
OS Homo sapiens.
XX
PN US2002001587-A1.
XX
PD 03-JAN-2002.
XX
PF 16-MAR-2001; 2001US-0811123.
XX
PR 16-MAR-2000; 2000US-189844P.
PR 05-OCT-2000; 2000US-238327P.
XX
PA (ERIC/) ERICKSON S.
PA (SCHW/) SCHWALL R.
PA (SLIW/) SLIWKOWSKI M.
XX
```

```
PI Erickson S, Schwall R, Sliwowski M;
XX WPI; 2002-163686/21.
DR N-PSDB; ABK14058.
XX
PT Treating tumour characterised by overexpression of epidermal growth
PT factor receptor, ErbB or cancer in mammal, comprises administering
PT anti-ErbB antibody-maytansinoid conjugate to the mammal -
XX
PS Example 3; Fig 7; 93pp; English.
XX
CC The invention relates to treating a tumour in a mammal, where the tumour
CC is characterised by the overexpression of an epidermal growth factor
CC receptor (ErbB) and does not respond or responds poorly, to treatment
CC with an anti-ErbB antibody, comprising administering to the mammal an
CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for
CC treating cancer or tumours of the breast, ovary, stomach, endometrium,
CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,
CC prostate and bladder, preferably breast cancer. The breast cancer is a
CC metastatic breast cancer or an aggressive form of metastatic breast
CC cancer which overexpresses ErbB2. The method is also useful for treating
CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,
CC epithelial, stromal, blastocoealic, inflammatory, angiogenic and
CC immunological disorders. This sequence represents the human HER2 (ErbB2)
CC polypeptide of the invention.
XX
SQ Sequence 1255 AA;
```

```
Query Match 98.1%; Score 6714; DB 23; Length 1255;
Best Local Similarity 98.3%; Pred. No. 0;
Matches 1243; Conservative 0; Mismatches 8; Indels 14; Gaps 2;

Qy 1 MELAALCRWGLLLALLPCCAASQVCTGTDMLKRLPASPEHLDMRLHLYQGCVQVGNL 60
Db 1 MELAALCRWGLLLALLPCCAASQVCTGTDMLKRLPASPEHLDMRLHLYQGCVQVGNL 60

Qy 61 ELTYLPTNASLFLQDIEVQGVLIHNRQVPLQRLIRVGTOLFDNYALAVLDNG 120
Db 61 ELTYLPTNASLFLQDIEVQGVLIHNRQVPLQRLIRVGTOLFDNYALAVLDNG 120

Qy 121 DFLNNFNFTVFWLVRPKVSAS-----HLEQLRSLTEILKGGVLIQRNPOLCYQDTTLWK 176
Db 121 DFLNN-----TTPVTGASPGGLRELRLSLTEILKGGVLIQRNPOLCYQDTTLWK 170

Qy 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKGRWCSESSEDCOSLRTVCAGGCARCKGP 236
Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKGRWCSESSEDCOSLRTVCAGGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRYT 290

Qy 297 FGASCVTACPNYVLTVDGSCITLVCPHNRQVTAEDGTORCEKCKSPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYVLTVDGSCITLVCPHNRQVTAEDGTORCEKCKSPCARVCYGLGMEHL 350

Qy 357 REVRAVTSANIOBFAGCKKIFGSLAPLPESFGDPPASNTAPLOPELOVFEETLEETGYL 416
Db 351 REVRAVTSANIOBFAGCKKIFGSLAPLPESFGDPPASNTAPLOPELOVFEETLEETGYL 410

Qy 417 YISAMPDSLPLDSVFNQLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRSLGSLALIH 476
Db 411 YISAMPDSLPLDSVFNQLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRSLGSLALIH 470

Qy 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVN 536
Db 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAACHOLCARGHCWGPGTQCVN 530

Qy 537 CSQFLRGQECVEECVRLQGLPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHY 596
Db 531 CSQFLRGQECVEECVRLQGLPREYNARHCLPCHPECPQNGSVTCFGEADOCVACAHY 590

Qy 597 KDPPFCVARCPGSKVPKPDLSYMPDIWPFDEGACQPCPINCTHSCVDLDDKGCFAEQRASP 656
```

Db	591	KDPFFCVCARCPGKVPDLSPYPIWKFPDEEGACQPCPINCTHSCVDLDDKGCPAEQRAS	650
Qy	657	LTISVSAVGIILLVWLVGVFGILIKRQOKIRKTYMRRLOETELVEPLTPSGAMPNOA	716
Db	651	LTISVSAVGIILLVWLVGVFGILIKRQOKIRKTYMRRLOETELVEPLTPSGAMPNOA	710
Qy	717	QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDE	776
Db	711	QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRRENTSPKANKEILDE	770
Qy	777	AYMAGVGSYVSRLLIGICTSTVQLVTQLMPYGCCLLDHVRENRGRIGSDOLLNWCQIA	836
Db	771	AYMAGVGSYVSRLLIGICTSTVQLVTQLMPYGCCLLDHVRENRGRIGSDOLLNWCQIA	830
Qy	837	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMA	896
Db	831	KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKIKMA	890
Qy	897	LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID	956
Db	891	LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID	950
Qy	957	VYIMVWCWIMIDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE	1016
Db	951	VYIMVWCWIMIDSECRPRFRELVSERFARMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE	1010
Qy	1017	DDMGDLVDAEYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLEPSEEA	1076
Db	1011	DDMGDLVDAEYLVPQGFPCPDPAFCAGGMVHRRSSSTRSGGDLTLGLEPSEEA	1070
Qy	1077	PRSPAPSEAGSDVDFDGLGMAAKGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVAP	1136
Db	1071	PRSPAPSEAGSDVDFDGLGMAAKGLQSLPTHDPSPLOKRYSEDPTVPLPSETDGYVAP	1130
Qy	1137	LTCSPOEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGA	1196
Db	1131	LTCSPOEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTILSPGKNGVVKDVFAFGA	1190
Qy	1197	VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYLG	1256
Db	1191	VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYLG	1250
Qy	1257	LDVPV 1261	
Db	1251	LDVPV 1255	
RESULT 5			
ID	AAW01111 standard; Protein; 1255 AA.		
XX			
AC	AAW01111;		
DT	01-JAN-1997 (first entry)		
XX			
DE	HER-2/neu protein.		
KW	HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;		
KW	breast cancer; ovary cancer; colon cancer; lung cancer;		
XX	prostate cancer; immunisation; tumour; vaccine; vector.		
OS	Homo sapiens.		
XX			
PH	Location/Qualifiers		
FT	676..1255		
FT	/label= intracellular domain		
XX	/note= "claimed domain, useful for immunisation"		
PN	MO9630514-A1.		
XX			
PD	03-OCT-1996.		
XX			

PF	28-MAR-1996;	96WO-US01689.	
XX	31-MAR-1995;	95US-0414417.	
PR	(UNIW ) UNIV WASHINGTON.		
PA	Cheever MA, Disais ML;		
XX	PI		
XX	WPI; 1996-455361/45.		
DR	N-PSDB; AAT40739.		
XX	DNA encoding HER-2-neu poly:peptide(s) - used for prevention or		
PT	treatment of malignancies with which the HER-2/neu oncogene is		
PT	associated		
XX	Claim 2; Page 56-61; 71pp; English.		
PS	Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is		
XX	the product of the HER-2/neu oncogene (see also AAT40739). The		
CC	protein is over-expressed in various cancers, including breast,		
CC	ovarian, colon, lung and prostate. The intracellular domain of the		
CC	protein can be used to immunise an animal against a malignancy with		
CC	which the oncogene is associated. The polypeptide can be produced		
CC	in transformed host cells for use in immunisation. Alternatively,		
CC	animal cells are transfected in vivo or ex vivo with a viral vector		
CC	that directs expression of the polypeptide.		
XX	Sequence 1255 AA;		
SQ	Query Match 98.0%; Score 6708; DB 17; Length 1255;		
	Best Local Similarity 98.1%; Pred. No. 0;		
	Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;		
Qy	1	MELAAALCRWGLLLALLPPGAASQVCTGDMKLRLPASPETHLDMLRHLVGGCQVQGNL	60
Db	1	MELAAALCRWGLLLALLPPGAASQVCTGDMKLRLPASPETHLDMLRHLVGGCQVQGNL	60
Qy	61	ELTYLPTNASLSFLQDIOEQVGYVLIAHNOVROVPLQRLRVIRGTQVLPEDNYALAVLNG	120
Db	61	ELTYLPTNASLSFLQDIOEQVGYVLIAHNOVROVPLQRLRVIRGTQVLPEDNYALAVLNG	120
Qy	121	DPLNNFNFTVSFWLRVPKVSAS----HLEQLRSLTEILKGGVLIQRNPQLCYQDTILWK	176
Db	121	DPLNN-----TTPVTGASPGGLRELQLRSLTEILKGGVLIQRNPQLCYQDTILWK	170
Qy	177	DI FHKNQALTLIDTNRBACHPCSPMCKSGRCWGESSEDCQSILTRTVCAAGGCARCKGP	236
Db	171	DI FHKNQALTLIDTNRBACHPCSPMCKSGRCWGESSEDCQSILTRTVCAAGGCARCKGP	230
Qy	237	LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT	296
Db	231	LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT	290
Qy	297	FGASCVTACPNYLSLTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHL	356
Db	291	FGASCVTACPNYLSLTDVSGCTLVCPHLNQEVTAEDGTQRCCKSKPCARVCYGLGMEHL	350
Qy	357	REVRVTSANIQSFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEITGYL	416
Db	351	REVRVTSANIQSFAGCKKIFGSLAFIPESFDGDPASNTAPLQPEQLQVFTLEITGYL	410
Qy	417	YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGSGLALIH	476
Db	411	YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTQGLGISWLGSLRSLRELGSGLALIH	470
Qy	477	NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVN	536
Db	471	NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVN	530
Qy	537	CSQFLRQECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQCVACAHY	596
Db	531	CSQFLRQECVBECEVRLQGLPREYVNAHCLPCHPECPQNGSVTCTGPEADQCVACAHY	590

QY 597 KDPPEFCVARGSGVXPDLSPYPIWKPPDEEGACQPCPINCTHSCVDLDDKGCFAEORASP 656  
DB |||||  
QY 591 KDPPEFCVARGSGVXPDLSPYPIWKPPDEEGACQPCPINCTHSCVDLDDKGCFAEORASP 650  
DB |||||  
QY 657 LTSISAVVGILLVVLVGVVGLIKRRQOKIRKVTMRLLQETELVEPLTPSGAMPNOA 716  
DB |||||  
QY 651 LTSISAVVGILLVVLVGVVGLIKRRQOKIRKVTMRLLQETELVEPLTPSGAMPNOA 710  
DB |||||  
QY 717 QMRILKETELRKVKVGLSGAGFTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDE 776  
DB |||||  
QY 711 QMRILKETELRKVKVGLSGAGFTVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDE 770  
DB |||||  
QY 777 AYVMAVGSPYSRLIGICLTSTVOLVTQMLPYGCLLDHVRENRLGSGQDILNWCMTA 836  
DB |||||  
QY 771 AYVMAVGSPYSRLIGICLTSTVOLVTQMLPYGCLLDHVRENRLGSGQDILNWCMTA 830  
DB |||||  
QY 837 KGMYSLEVDRLVHROLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 896  
DB |||||  
QY 831 KGMYSLEVDRLVHROLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 890  
DB |||||  
QY 897 LESILRRPETHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956  
DB |||||  
QY 891 LESILRRPETHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950  
DB |||||  
QY 957 VYIMVWKWIDSECRPRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1016  
DB |||||  
QY 951 VYIMVWKWIDSECRPRELVSFSEMRARDPQRFVVIQNEIDLGPASPLDSTFYRSLE 1010  
DB |||||  
QY 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1076  
DB |||||  
QY 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGMVHRRSSSTRSGGDLTLGLEPSEEA 1070  
DB |||||  
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAP 1136  
DB |||||  
QY 1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPVLPSETDGYVAP 1130  
DB |||||  
QY 1137 LTCSPQEVNQPDRPPOPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGGA 1196  
DB |||||  
QY 1131 LTCSPQEVNQPDRPPOPPSPREGPLPAARAGATLERAKTSLPGKNGVVKDVFAGGA 1190  
DB |||||  
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256  
DB |||||  
QY 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250  
DB |||||  
QY 1257 LDVPV 1261  
DB |||||  
QY 1251 LDVPV 1255  
DB |||||

RESULT 6

AAW92406  
ID AAW92406 standard; Protein; 1255 AA.

AC AAW92406;

DT 21-APR-1999 (first entry)

XX Human HER-2/neu oncogene protein.

DE HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
KW malignancy; treatment; tumour.

OS Homo sapiens.

XX Key Location/Qualifiers

FH 676..1255

FT /note= "region which elicits immune response"

XX US869445-A.

PN 09-FEB-1999.

PD 01-APR-1996; 96US-0625101.

XX

PF

XX 01-APR-1996; 96US-0625101.  
PR 17-MAR-1993; 93US-0033644.  
PR 12-AUG-1993; 93US-0106112.  
PR 31-MAR-1995; 95US-0414417.  
PA (UNIW ) UNIV WASHINGTON.  
XX  
XX Cheever MA, Disis ML;  
PI WPI: 1999-152835/13.  
DR N-PSDB; AAX01912.  
XX  
PT Use of HER-2/neu polypeptides - for eliciting an immune response to  
PT an HER-2/neu associated malignancy, particularly for treating or  
PT preventing tumours  
XX  
PS Claim 3; Column 31-38; 26pp; English.  
XX  
CC This sequence represents the human HER-2/neu oncogene protein. A fragment  
CC of this protein is used in a method for eliciting or enhancing an immune  
CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
CC B cells to produce an immune response to the HER-2/neu protein. The  
CC method can be used for immunisation against a malignancy in which the  
CC HER-2/neu oncogene is associated and in the treatment of an existing  
CC tumour, or to prevent tumour occurrence or reoccurrence.  
XX  
SQ Sequence 1255 AA;  
Query Match 98.0%; Score 6708; DB 20; Length 1255;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;  
QY 1 MELAALCRWGLLLALLPGGAASQVCTGDMKRLPASPETHLDMLRHLYQGCQVVOGNNL 60  
DB |||||  
QY 1 MELAALCRWGLLLALLPGGAASQVCTGDMKRLPASPETHLDMLRHLYQGCQVVOGNNL 60  
DB |||||  
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNVQVPLQRLRIRVGTQOLFEDNALAVLDNG 120  
DB |||||  
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNVQVPLQRLRIRVGTQOLFEDNALAVLDNG 120  
DB |||||  
QY 121 DPLNNFNFTVSFWLRVPKVSAS----HLEQLRSITEILKGGVLIQRNPOLCYQDTILWK 176  
DB |||||  
QY 121 DPLNNFNFTVSFWLRVPKVSAS----HLEQLRSITEILKGGVLIQRNPOLCYQDTILWK 170  
DB |||||  
QY 177 DIFHKNNQALTLIDTNESRACHPCSPMKSGRCSESEDCOSLTRTVCAGGCARCKGP 236  
DB |||||  
QY 171 DIFHKNNQALTLIDTNESRACHPCSPMKSGRCSESEDCOSLTRTVCAGGCARCKGP 230  
DB |||||  
QY 237 LPTDCCHQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296  
DB |||||  
QY 231 LPTDCCHQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290  
DB |||||  
QY 297 FGASCVTACPNYLSLTDVGSCTLVCPLNHNVQVTAEDGTORCEKSKPCARVCYGLMEHL 356  
DB |||||  
QY 291 FGASCVTACPNYLSLTDVGSCTLVCPLNHNVQVTAEDGTORCEKSKPCARVCYGLMEHL 350  
DB |||||  
QY 357 REVRAVTSANTQEPAGCKKIFGSLAPLPESDPASNTAPLOPQLOVFEETLEETGYL 416  
DB |||||  
QY 351 REVRAVTSANTQEPAGCKKIFGSLAPLPESDPASNTAPLOPQLOVFEETLEETGYL 410  
DB |||||  
QY 417 YISAMPDLSPLDSVFQNLQVIRGRILHNGAYSLTLQGLGISMLGLRSRELGLALIIH 476  
DB |||||  
QY 411 YISAMPDLSPLDSVFQNLQVIRGRILHNGAYSLTLQGLGISMLGLRSRELGLALIIH 470  
DB |||||  
QY 477 NTHLCFVHTVPWDQLFRPNHQAALLHTANRPEDECVGEGLAGHQLCARGHCWGPGTQCVN 536  
DB |||||  
QY 471 NTHLCFVHTVPWDQLFRPNHQAALLHTANRPEDECVGEGLAGHQLCARGHCWGPGTQCVN 530  
DB |||||  
QY 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCACAHY 596  
DB |||||  
QY 531 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECOPOGNSVTCFGEADQCACAHY 590  
DB |||||

QY 597 KDPFFCVARCPGVKPDLSYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGPAPQASRP 656  
Db 591 KDPFFCVARCPGVKPDLSYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGPAPQASRP 650  
QY 657 LTSIVSAGVGIILVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOA 716  
Db 651 LTSIIISAVGIIILVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOA 710  
QY 717 QNRILKETELRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDE 776  
Db 711 QNRILKETELRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRENTSPKANKIILDE 770  
QY 777 AYVMAGVSPYVSRLIGICLTSTVQLVTQMLPVGCLLDHVRENRLGSGQDLLLLNWCQIA 836  
Db 771 AYVMAGVSPYVSRLIGICLTSTVQLVTQMLPVGCLLDHVRENRLGSGQDLLLLNWCQIA 830  
QY 837 KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDDIDETEHADGGKVPKWA 896  
Db 831 KGMSYLEDVRLVHRDLAARNVLKSPNHVKITDGLARLLDDIDETEHADGGKVPKWA 890  
QY 897 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956  
Db 891 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950  
QY 957 VYIMVYKWMIDSECRPRFRELVSFSEFMRARDPQRFVJQNEDLGPASPLDSTFYRSLE 1016  
Db 951 VYIMVYKWMIDSECRPRFRELVSFSEFMRARDPQRFVJQNEDLGPASPLDSTFYRSLE 1010  
QY 1017 DDDMGDLVDAEYLVLPQGGFFCDDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEA 1076  
Db 1011 DDDMGDLVDAEYLVLPQGGFFCDDPAPGAGGMVHHRSSSTRSGGDLTLGLEPSEEA 1070  
QY 1077 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136  
Db 1071 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130  
QY 1137 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVYKDVAFGGA 1196  
Db 1131 LTCSPQPEYVNOPDVRPQPPSPREGPLPAARPAAGATLERAKTLPSPKNGVYKDVAFGGA 1190  
QY 1197 VENPEYLTQGGGAOPHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256  
Db 1191 VENPEYLTQGGGAOPHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250  
QY 1257 LDVPV 1261  
Db 1251 LDVPV 1255

RESULT 7  
ID AAB21198  
XX AAB21198 standard; protein; 1255 AA.  
AC AAB21198;  
XX 12-JAN-2001 (first entry)  
DE Human HER-2/neu protein.  
DE Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;  
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;  
KW colon cancer.  
XX Homo sapiens.  
XX WO200044899-A1.  
XX 03-AUG-2000.  
XX 28-JAN-2000; 2000WO-US02164.  
PF 29-JAN-1999; 99US-0117976.  
PR  
XX

PA (CORI-) CORIXA CORP.  
PA (SMIK) SMITHKLINE BEECHAM.  
PI Cheever MA, Gheysen D;  
XX WPI; 2000-505976/45.  
DR N-PSDB; AAA89736.  
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins  
PT useful for vaccinating against breast, ovarian, colon, lung and  
PT prostate cancers -  
XX Claim 52; Fig 7; 128pp; English.  
PS The present sequence is the human HER-2/neu protein. It is a member of  
CC the tyrosine kinase family of receptor-like glycoproteins and shows  
CC homology to the epidermal growth factor receptor (EGFR). It probably  
CC plays a part in cell growth and/or differentiation. The HER-2/neu  
CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
CC domain may be used to treat or prevent cancer by eliciting or  
CC enhancing an immune response to the HER-2/neu protein. It may be used  
CC to treat malignancies such as breast, ovarian, colon, lung and  
CC prostate cancers, and may be used as an antigen to vaccinate against  
CC these neoplasias.  
XX SQ Sequence . 1255 AA;

Query Match 98.0%; Score 6708; DB 21; Length 1255;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;  
QY 1 MELAALCRWGLLALLPGCAASTQVCTGDMKRLRLPASPTHLDMLRHLVQGCQVQGNL 60  
Db 1 MELAALCRWGLLALLPGCAASTQVCTGDMKRLRLPASPTHLDMLRHLVQGCQVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIEVQGVYLIAHNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIEVQGVYLIAHNQVQVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
QY 121 DFLNNFNFTVFWLRVPKVSAS----HLEQLRSLTEILKGGVLIQORNQOLCYQDITLWK 176  
Db 121 DFLNN-----TTPVTGASPGGLRELRLSLTEILKGGVLIQORNQOLCYQDITLWK 170  
QY 177 DIFPHKNQALTLIDTNRSRACHPCSPCKSGRCWSESSEDQSLTRTVCCAGCARCKGP 236  
Db 171 DIFPHKNQALTLIDTNRSRACHPCSPCKSGRCWSESSEDQSLTRTVCCAGCARCKGP 230  
QY 237 LPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFSMPNPEGRT 296  
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTTFSMPNPEGRT 290  
QY 297 FGASCVTACPNYVLTVDVGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCVGLGMEHL 356  
Db 291 FGASCVTACPNYVLTVDVGSCTLVCPHLNQEVTAEDGTORCEKSPCARVCVGLGMEHL 350  
QY 357 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDPPASNTAPLOPEQLQVFTLEEITGYL 416  
Db 351 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDPPASNTAPLOPEQLQVFTLEEITGYL 410  
QY 417 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIH 476  
Db 411 YISAWPDSLPLDSVFQNLQVIRGRILHNGAYSILTLQGLGISWLGRLSLRELGSGLALIH 470  
QY 477 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPGPTQCVN 536  
Db 471 NTHLCFVHTVPWDQLFRNPHQALLHTANRPEDECVGEGLAACHQLCARGHCWGPGPTQCVN 530  
QY 537 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 596  
Db 531 CSQFLRGQECVEECRVLQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHY 590  
QY 597 KDPFFCVARCPGVKPDLSYMPIWKFPPDEEGACQPCPINCTHSCVDLDDKGPAPQASRP 656

Db 591 KDPFFCVARCPGKPDLSMP1WKFPDEEGACQPCPINCTHSCVDLDDKCPAEQRASP 650  
Qy 657 LTSIVSAVVGILLVWLVGVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA 716  
Db 651 LTSIIISAVVGILLVWLVGVFGILIKRQOKIRKYTMRRLLQETELVEPLTPSGAMPNQA 710  
Qy 717 QMRILKETELRKVKVLGSGAFGTYYGIWIIPDGENVKIPVAIKVLRNTPSPKANKETLDE 776  
Db 711 QMRILKETELRKVKVLGSGAFGTYYGIWIIPDGENVKIPVAIKVLRNTPSPKANKETLDE 770  
Qy 777 AYVMAGVGSPPVSRLLGICLTSTVOLVTQMLPVGCLLDHVRENRLGSDODLNNWCQIA 836  
Db 771 AYVMAGVGSPPVSRLLGICLTSTVOLVTQMLPVGCLLDHVRENRLGSDODLNNWCQIA 830  
Qy 837 KGMYSLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWWA 896  
Db 831 KGMYSLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKGVPIKWWA 890  
Qy 897 LESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTID 956  
Db 891 LESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPOLLEKGERLPPOPICTID 950  
Qy 957 VYIMVWKWIMIDSECRPRFRELSEFSRMDRQRFVVIQNEDLGPASPLDSTFYRSLLE 1016  
Db 951 VYIMVWKWIMIDSECRPRFRELSEFSRMDRQRFVVIQNEDLGPASPLDSTFYRSLLE 1010  
Qy 1017 DDDMGDLVDAEYLVFQQGFCDPAPAGMVGHHRRSSSTRSGGDLTLGLEPSEEEA 1076  
Db 1011 DDDMGDLVDAEYLVFQQGFCDPAPAGMVGHHRRSSSTRSGGDLTLGLEPSEEEA 1070  
Qy 1077 PRSPLAPSGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDVPLPSETDGYVAP 1136  
Db 1071 PRSPLAPSGAGSDVFDGLGMAAKGLOSPLTHDPSPLQRYSEDVPLPSETDGYVAP 1130  
Qy 1137 LTCSPQPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGA 1196  
Db 1131 LTCSPQPEYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTSLPGKNGVVKDVFAFGGA 1190  
Qy 1197 VENPEYLTQGGGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256  
Db 1191 VENPEYLTQGGGAAPQHPHPPAPSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250  
Qy 1257 LDVPV 1261  
Db 1251 LDVPV 1255

RESULT 8

AA84780  
ID AA84780 standard; Protein; 1255 AA.

XX AA84780;

AC AA84780;

DT 08-AUG-2000 (first entry)

DE Amino acid sequence of the SPLICE erbb-2 receptor protein.

XX SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;  
KW tumor cell proliferation; tissue degeneration; arthropathy;  
KW bone resorption; inflammatory disease; degenerative disorder;  
KW wound healing.

OS Homo sapiens.

XX WO200020579-A1.

PN 13-APR-2000.

PD 01-OCT-1999; 99WO-CA00912.

PF 02-OCT-1998; 98US-0165192.

PR 02-OCT-1998; 98US-0165192.

XX 02-OCT-1998; 98US-0165192.

(UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

PI WPI: 2000-303768/26.

DR N-PSDB; AAA14812.

XX Nucleic acid encoding an erbb 2 receptor protein designated SPLICE

PT erbb-2, inhibitors of the protein are useful for treatment of cancer -

XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPLICE erbb-2 receptor protein. The

CC protein has an in-frame deletion of 16 amino acids, 2 of which are

CC conserved cysteine residues, compared to the unspliced protein. The

CC erbb-2 polynucleotide is used to construct probes for detecting

CC disorders of cell transformation such as cancer. Antibodies to the

CC protein may be used to detect SPLICE erbb-2 in a sample. Agents

CC (e.g. antisense oligonucleotides) which inhibit the expression of

CC SPLICE erbb-2 are useful for reducing tumor cell proliferation and

CC treating cancer. Substances which stimulate SPLICE erbb-2 are useful

CC for treating conditions involving damaged cells including conditions

CC in which degeneration of tissue occurs, such as arthropathy, bone

CC resorption, inflammatory diseases, degenerative disorders of the

CC central nervous system and wound healing.

XX SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 21; Length 1255;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

Qy 1 MELAALCRWGLLIALLPGAASQVCTGTDKMLRLPASPTHLDMLRHLVQGCVOVQGNL 60

Db 1 MELAALCRWGLLIALLPGAASQVCTGTDKMLRLPASPTHLDMLRHLVQGCVOVQGNL 60

Qy 61 ELTYLPTNASLSPLQIQEVQGVYLIHNVQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Db 61 ELTYLPTNASLSPLQIQEVQGVYLIHNVQVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120

Qy 121 DPLNNFNTVSPWLRVVKVAS-----HLEQLRSLTEILKGGVLIQRNPOLCYQDTTLWK 176

Db 121 DPLNNFNTVSPWLRVVKVAS-----HLEQLRSLTEILKGGVLIQRNPOLCYQDTTLWK 170

Qy 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWGESSEDCOSLTRTVCAGGCARCKGP 236

Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGCWGESSEDCOSLTRTVCAGGCARCKGP 230

Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296

Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290

Qy 297 FGASCVTACPNYVLTVDGSCITLVCPHNOEVTAEQTCRCEKSPCARVCYGLGMEHL 356

Db 291 FGASCVTACPNYVLTVDGSCITLVCPHNOEVTAEQTCRCEKSPCARVCYGLGMEHL 350

Qy 357 REVRAVTSANIOEFAGCKIFGSLAPLPSFDGDPASNTAPLOEQLOVFTLEETIGYL 416

Db 351 REVRAVTSANIOEFAGCKIFGSLAPLPSFDGDPASNTAPLOEQLOVFTLEETIGYL 410

Qy 417 YISAWPDSLPLDSVFNQNLQVIRGRILHNGAYSLLTQGLGISMWGLRSLRSLGSLALIH 476

Db 411 YISAWPDSLPLDSVFNQNLQVIRGRILHNGAYSLLTQGLGISMWGLRSLRSLGSLALIH 470

Qy 477 NTHLCFVHTVPWDOLFNRPHQALLHTANRPEDECVEGEGACHQLCARGHCWGPGTQCVN 536

Db 471 NTHLCFVHTVPWDOLFNRPHQALLHTANRPEDECVEGEGACHQLCARGHCWGPGTQCVN 530

Qy 537 CSQFLRGQCEVEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHY 596

Db 531 CSQFLRGQCEVEECRVLQGLPREYVNAHCLPCHPECCQPNQSVTCFGEADQCVACAHY 590

Qy 597 KDPPFCVARCPGSKVPDLSYMP1WKFPDEEGACQPCPINCTHSCVDLDDKCPAEQRASP 656

Db	591	KDPFPCVARGCSGVKPLDSYMP1WKFPDEGACQPCPINCTHSCVDLDDKGCFAEQRAS	650
Qy	657	LTSVSAVVGILLVVVLGVVFGILLIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQA	716
Db	651	LTSIISAVVGILLVVVLGVVFGILLIKRRQKIRKYTMRRLLQETELVEPLTPSGAMPNQA	710
Qy	717	QMRILKETELURKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELLDE	776
Db	711	QMRILKETELURKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKELLDE	770
Qy	777	AYVMAGVGSPPVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGODLLNWCMOIA	836
Db	771	AYVMAGVGSPPVSRLLGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSGODLLNWCMOIA	830
Qy	837	KGMSYLEBVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP1KWWA	896
Db	831	KGMSYLEBVLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVP1KWWA	890
Qy	897	LESLRRRFTHOSDVWISYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP1CTID	956
Db	891	LESLRRRFTHOSDVWISYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPP1CTID	950
Qy	957	YYIMVWKCMWIDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLL	1016
Db	951	YYIMVWKCMWIDSECRPRFRELVSSEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLL	1010
Qy	1017	DDMGDLVDAEYILVPOGFFCPDPAPGAGMWHHRSSSTRSGGGDLTLGLEPSEEEA	1076
Db	1011	DDMGDLVDAEYILVPOGFFCPDPAPGAGMWHHRSSSTRSGGGDLTLGLEPSEEEA	1070
Qy	1077	PRSLAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLOYSIEDPTVPLPSETDGYVAP	1136
Db	1071	PRSLAPSEAGSDVFDGDLGMAAKGLQSLPTHDPSPLOYSIEDPTVPLPSETDGYVAP	1130
Qy	1137	LTCSPQPEYVNPQDVRQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAPGGA	1196
Db	1131	LTCSPQPEYVNPQDVRQPPSPREGPLPAARPAGATLERAKTILSPGKNGVVKDVFAPGGA	1190
Qy	1197	VENPEYILTPOGGAAPQPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG	1256
Db	1191	VENPEYILTPOGGAAPQPHPPAPSPAFDNLVYWDQDPPERGAPPSTFKGTPTAENPEYLG	1250
Qy	1257	LDVPV 1261	
Db	1251	LDVPV 1255	
RESULT 9			
AAB85458			
ID	AAB85458	standard; Protein; 1255 AA.	
XX	XX		
AC	AAB85458;		
XX	XX		
DT	25-SEP-2001	(first entry)	
XX	XX		
DE	Human HER-2/neu protein.		
DE	XX		
XX	XX		
KW	KW	Antigen-presenting cell; immunogenic; immune response; HER-2/neu;	
KW	KW	oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.	
XX	XX		
OS	XX	Homo sapiens.	
XX	XX		
FN	XX	WO200153463-A2.	
XX	XX		
PD	XX	26-JUL-2001.	
XX	XX		
PF	XX	19-JAN-2001; 2001WO-US01850.	
XX	XX		
PR	XX	21-JAN-2000; 2000US-0177545.	
XX	XX		
XX	XX	(CORI-) CORIXA CORP.	
XX	XX		

```
Db 651 LTSIIISAVVGIIILVVVLGVVFGILIKRQOKIRKYTMRRLLQETELVELTSGAMPNOA 710
Qy 717 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPRANKIILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPRANKIILDE 770
Qy 777 AYVMAGVGSYVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSQDILLNWCQIA 836
Db 771 AYVMAGVGSYVSRLLIGICLTSTVQLVTQMLPYGCLLDHVRNRRGLSQDILLNWCQIA 830
Qy 837 KGMYSLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 896
Db 831 KGMYSLEVDRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 890
Qy 897 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 956
Db 891 LESILRRRTHQSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPPICTID 950
Qy 957 VYMIWVKMIDSECRPRELVSERMRDQRFVWIONEDLGPASPLDSTFYRSLLE 1016
Db 951 VYMIWVKMIDSECRPRELVSERMRDQRFVWIONEDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGVMVHRHRSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVPOQGFPCDPAPGAGGVMVHRHRSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVVKDVFAGGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARAGATLERAKTSLSPGKNGVVKDVFAGGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPRSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDODPPERGAPRSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 10
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
XX
XX Keogh E;
XX
XX WPI; 2001-374995/39.
```

```
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX Disclosure; Page 15; 199pp; English.
XX
CC The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding (II); (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample from a patient and
CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected epitopes may be combined to
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.
XX
SQ Sequence 1255 AA;
```

```
Query Match 98.0%; Score 6708; DB 22; Length 1255;
Best Local Similarity 98.1%; Pred. No. 0;
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

Qy 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYOGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLYOGCVVQGNL 60
Qy 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIOEVQGYVLIAHNQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNFNFTVSFWLRVPKVSAS----HLEQLRSLTEILKGVLIQRPOLCYODTILWK 176
Db 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILWK 170
Qy 177 DIFHKNNQALTLIDTNRSRACHPCSPMKGSRCHGSESDCSQSLTRTYCAGGCARCKGP 236
Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMKGSRCHGSESDCSQSLTRTYCAGGCARCKGP 230
Qy 237 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESWPNPGRYT 296
Db 231 LPTDCCHQCAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFESWPNPGRYT 290
Qy 297 FGASCVTACPNYLLSTDVGSCTLVCPLNHQNVEAEDGTQRCCKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLLSTDVGSCTLVCPLNHQNVEAEDGTQRCCKSKPCARVCYGLGMEHL 350
Qy 357 REVRAVTSANTQEFAGCKKIFGSLAFLESFGDPSANTAPLOPEOLQVFTEETGYL 416
Db 351 REVRAVTSANTQEFAGCKKIFGSLAFLESFGDPSANTAPLOPEOLQVFTEETGYL 410
Qy 417 YISAWPDSLPLDSVFQNLQVIRGRITLHNGAVSLTLQGLGISWLGHSRRELGSGLAIHH 476
Db 411 YISAWPDSLPLDSVFQNLQVIRGRITLHNGAVSLTLQGLGISWLGHSRRELGSGLAIHH 470
Qy 477 NTHLCFVHTVPMQDLFRNPQHALLHTANRPEDCVGEGGLACHQLCARGHCWGPQTQCVN 536
Db 477 NTHLCFVHTVPMQDLFRNPQHALLHTANRPEDCVGEGGLACHQLCARGHCWGPQTQCVN 536
```



```
Db 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRDEBCEVGEGLACHOLCARGHCWGPGTQCVN 530
Qy 537 CSOFLRQECVECRVLQGLPREYVVARHCLPCHPECOPOGVSVTFCGPEADOCVACAHY 596
Db 531 CSOFLRQECVECRVLQGLPREYVVARHCLPCHPECOPOGVSVTFCGPEADOCVACAHY 590
Qy 597 KDPFFCVARCPGKVPDLSYMPYKFPDEEGACQPCPCINCHSCVDLDDKGCPCAEQASAP 656
Db 591 KDPFFCVARCPGKVPDLSYMPYKFPDEEGACQPCPCINCHSCVDLDDKGCPCAEQASAP 650
Qy 657 LTSIVANVGILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOA 716
Db 651 LTSIIISAVVIGILLVWLVGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOA 710
Qy 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDE 776
Db 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSKANKETLDE 770
Qy 777 AYVMAGVGSYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIA 836
Db 771 AYVMAGVGSYVSRLLGICLTSTVQLVTQMLPYGCLLDHVRENRLGSDLLNWCQIA 830
Qy 837 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 896
Db 831 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950
Qy 957 VYIMVCKMIMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1016
Db 951 VYIMVCKMIMIDSECRPRFRELVSFMRMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDMGDLVDAEYLVPOQGFCDPAPGAGMWHHRSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPLTHDPSPLQRYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPLTHDPSPLQRYSEDPVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKGVVXKDVAFPGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTLSPGKGVVXKDVAFPGA 1190
Qy 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERCAPPSTFKGTPTAENPYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLYYWDQDPPERCAPPSTFKGTPTAENPYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255
```

## RESULT 11

AAE24067

ID AAE24067 standard; Protein; 1255 AA.

XX

AC AAE24067;

XX

DT 23-SEP-2002 (first entry)

XX

DE Human Her-2 protein.

XX

KW Human; Her-2; epidermal growth factor receptor 2; infection; cancer;

KW hyperproliferative disorder; prophylaxis; inflammation; antisense;

KW tumour; gene therapy; phosphorothioate backbone.

XX

OS Homo sapiens.

XX

PN W020222636-A1.

XX

```
PD 21-MAR-2002.
XX
PF 12-SEP-2001; 2001WO-US28572.
XX
PR 15-SEP-2000; 2000US-0663834.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Bennett CF, Cowbert LM;
XX
XX WPI; 2002-471192/50.
DR N-PSDB; AAD38904.
XX
XX Novel antisense oligonucleotide which modulates the expression of Human
PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT inflammation or to prevent infection in humans -
XX
XX Example 13; Page 95-107; 116pp; English.
XX
XX The invention relates to antisense compounds targetted to a nucleic
CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)
CC that specifically hybridises with and inhibits the expression of Her2.
CC Antisense compounds of the invention are used for treating diseases or
CC conditions associated with Her2 such as hyperproliferative disorders
CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,
CC neural or cardiac cancer. They are also useful prophylactically e.g.
CC to prevent or delay infection, inflammation and tumour formation. The
CC invention is also used in gene therapy. The present sequence is human
CC Her-2 protein.
XX
XX Sequence 1255 AA;
```

```
Query Match 98.0%; Score 6708; DB 23; Length 1255;
Best Local Similarity 98.1%; Pred No. 0;
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;
```

```
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPEHLMLRLHLYGCGQVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLKRLPASPEHLMLRLHLYGCGQVQGNL 60
Qy 61 ELTYLPNASLSFLDIOEQVGVYVLAHNOVQVPLQRLIRVGTQTFEDNYALAVLDNG 120
Db 61 ELTYLPNASLSFLDIOEQVGVYVLAHNOVQVPLQRLIRVGTQTFEDNYALAVLDNG 120
Qy 121 DPLNNFNFTVSWLVRPKVSAS- ---HLEQLASLTETLKGVLQIQRNPOLCYODTTLWK 176
Db 121 DPLNN- ---TTPTVGASPGGURELQLSLTETLKGVLQIQRNPOLCYODTTLWK 170
Qy 177 DIFHKNQLALTLDITNRSRACHPCSPMKGSRGSGESSEDCQSLTRTVCAAGGCARCKGP 236
Db 171 DIFHKNQLALTLDITNRSRACHPCSPMKGSRGSGESSEDCQSLTRTVCAAGGCARCKGP 230
Qy 237 LPTDCCHEQCAAGCTGPKHSDCLACLFNHSIGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCCHEQCAAGCTGPKHSDCLACLFNHSIGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPNYLTSTDVSGCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYLTSTDVSGCTLVCPLNHNEVTAEDGTQRCCKSKPCARVCYGLGMEHL 350
Qy 357 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETGYL 416
Db 351 REVRVTSANIQEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEQLQVFTLEETGYL 410
Qy 417 YISAWPDSLPLSVFQNLQVIRGRILHNAGYSLTQLGIGISWLGSLRLSRLGSLALIH 476
Db 411 YISAWPDSLPLSVFQNLQVIRGRILHNAGYSLTQLGIGISWLGSLRLSRLGSLALIH 470
Qy 477 NTHLCFVHTVPMDQLFRNPHQALLHTANRDEBCEVGEGLACHOLCARGHCWGPGTQCVN 536
Db 471 NTHLCFVHTVPMDQLFRNPHQALLHTANRDEBCEVGEGLACHOLCARGHCWGPGTQCVN 530
Qy 537 CSOFLRQECVECRVLQGLPREYVVARHCLPCHPECOPOGVSVTFCGPEADOCVACAHY 596
```

Db 531 CSOFLRGQECVECRVQLGPREYVNRHCLPCHPECQPNQSVTCFGEADQCACAHY 590  
QY 597 KPPFFCVARCPGKVPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEQRASP 656  
Db 591 KPPFFCVARCPGKVPDLSPYMPKPFDEEGACQPCPINCTHSCVDLDDKGCFAEQRASP 650  
QY 657 LTSIVSAVVGILLVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOA 716  
Db 651 LTSIIISAVVGILLVVLGVVFGILIKRQOKIRKYMRRLLQETELVEPLTPSGAMPNOA 710  
QY 717 QMRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDE 776  
Db 711 QMRILKETELRKVKVLSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPSPKANKILDE 770  
QY 777 AYVMAGVSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDQLLNCMQIA 836  
Db 771 AYVMAGVSPYVSRLLGICLTSTVQLVTLQMPYGCCLLDHVRENRLGSGDQLLNCMQIA 830  
QY 837 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 896  
Db 831 KMSYLEDLVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKMA 890  
QY 897 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 956  
Db 891 LESILRRRTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTID 950  
QY 957 VYIMVYKMWIDSECRPRFRELVSERFMRDQRFVWQNEIDLGSPASPLDSTFYRSLLE 1016  
Db 951 VYIMVYKMWIDSECRPRFRELVSERFMRDQRFVWQNEIDLGSPASPLDSTFYRSLLE 1010  
QY 1017 DDMGDLVDAEYLVPOQGFCDPDPAPGAGGVMVHRRSSSTRSGGDLTLGLEPSEEA 1076  
Db 1011 DDMGDLVDAEYLVPOQGFCDPDPAPGAGGVMVHRRSSSTRSGGDLTLGLEPSEEA 1070  
QY 1077 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVLPSETDGYVAP 1136  
Db 1071 PRSPLAPSGAGSDVFDGLGMAAGLQSLPHTDPSPLQRYSEDPTVLPSETDGYVAP 1130  
QY 1137 LTCSPQEVYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVYKDVAFGGA 1196  
Db 1131 LTCSPQEVYVNOVDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVYKDVAFGGA 1190  
QY 1197 VENPEYLTQCGNAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLG 1256  
Db 1191 VENPEYLTQCGNAPQHPHPPAFSPAFDNLVYWDQPPERGAPPSTFKGTPTAENPEYLG 1250  
QY 1257 LDVPV 1261  
Db 1251 LDVPV 1255  
RESULT 12  
ID AAE20479  
XX AAE20479 standard; Protein; 1255 AA.  
AC AAE20479;  
XX 01-JUL-2002 (first entry)  
DT Human Her-2/neu protein.  
XX Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
KW human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
XX Homo sapiens.  
OS  
FH Key Location/Qualifiers  
FT 1021..1030  
FT /note= "Naturally processed HLA-B44-restricted epitope"  
XX  
PN WO200214503-A2.  
XX

PD 21-FEB-2002.  
XX 14-AUG-2001; 2001WO-US41733.  
XX 14-AUG-2000; 2000US-225152P.  
PR 28-SEP-2000; 2000US-236428P.  
PR 21-FEB-2001; 2001US-270520P.  
XX (CORI-) CORIXA CORP.  
XX Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;  
PI McNeill PD, Vedvick TS;  
XX WPI; 2002-280758/32.  
DR N-PSDB; AAD32743.  
XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
PT prevention and diagnosis of cancer, preferably breast cancer -  
XX Disclosure; Page 114-117; 129pp; English.  
XX The invention relates to an isolated Her-2/Neu polypeptide composition  
CC effective for eliciting an immune response. The invention is useful for  
CC eliciting an immune response in a patient, where the patient is human  
CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
CC The composition is useful for the therapy and diagnosis of cancer,  
CC preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
CC and other compositions for the diagnosis, prevention and treatment of  
CC human malignancies, for stimulating and/or expanding T cells specific for  
CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
CC patient. The invention is useful for stimulating a T cell response in a  
CC human patient, as probe or primer for nucleic acid hybridisation, to  
CC selectively form duplex molecules with complementary stretches of the  
CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
CC length gene from a suitable library, and to direct expression of a  
CC polypeptide in appropriate host cells. The composition is useful in  
CC prophylactic or therapeutic applications and for the treatment of cancer,  
CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
CC associated malignancies. The invention is useful in gene therapy. The  
CC present sequence is human Her-2/neu protein.  
XX Sequence 1255 AA;  
QY Query Match 98.0%; Score 6708; DB 23; Length 1255;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;  
Db 1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRLHLYQGCVVQGNL 60  
1 MELAALCRWGLLLALLPPGAASQVCTGDMKRLPASPETHLDMLRLHLYQGCVVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIEQVGVLIHQNQVPLQRLRIRVGTOLFEDNYALAVLDNG 120  
Db 61 ELTYLPTNASLSFLQDIEQVGVLIHQNQVPLQRLRIRVGTOLFEDNYALAVLDNG 120  
QY 121 DPLNNFNFTVSFWRVLPKVSAS----HLEQLRSLTEILKGGVLIQRPOLCYQDTILWK 176  
121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQRPOLCYQDTILWK 170  
QY 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGCARCKGP 236  
Db 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKSGRCWGESSEDCOSLTRTVCAGCARCKGP 230  
QY 237 LPTDCCHEOCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMNPGRY 296  
231 LPTDCCHEOCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMNPGRY 290  
QY 297 FGASCVTACPNYLSLDVSGSCTLVCPHNRQEVTAEDGTORCEKSKPCARVCYGLGMEHL 356  
291 FGASCVTACPNYLSLDVSGSCTLVCPHNRQEVTAEDGTORCEKSKPCARVCYGLGMEHL 350  
QY 357 REVRAVTSANIOEPAGCKKIFGSLAPLPESFGDGPASNTAPLOEQVFTLEETIGYL 416  
|||||

```
Db 351 REVRVTSANIQEFAGCKIFGSLAFLPESFDGDPASNTAPLQEQVFTLEBITGYL 410
Qy 417 YISAWPDSLPLSVFQNLQVIRGRI LHNAGYSLTLOGIGISWICLSRLSELGSLAIHH 476
Db 411 YISAWPDSLPLSVFQNLQVIRGRI LHNAGYSLTLOGIGISWICLSRLSELGSLAIHH 470
Qy 477 NTHLCFVHTVPWDLFRNPHQALLHTANRDECEVCGEGLACHQLCARGHCWGPOPTQCVN 536
Db 471 NTHLCFVHTVPWDLFRNPHQALLHTANRDECEVCGEGLACHQLCARGHCWGPOPTQCVN 530
Qy 537 CSQFLRQCEVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFPGPADQCAVCAHY 596
Db 531 CSQFLRQCEVEECRVLQGLPREYVYARHCLPCHPECPQNGSVTCFPGPADQCAVCAHY 590
Qy 597 KDPFFCVARCPGKVPDLSTYMPITWKFPEDEGACQPCPINTCHSCVDLDDKCPAEORASP 656
Db 591 KDPFFCVARCPGKVPDLSTYMPITWKFPEDEGACQPCPINTCHSCVDLDDKCPAEORASP 650
Qy 716 LTSIVSAVVGILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOA 716
Db 651 LTSIVSAVVGILLVVLGVVFGILIKRRQOKIRKYTMRRLLQETELVEPLTPSGAMPNOA 710
Qy 776 QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKETILDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRENTSPKANKETILDE 770
Qy 836 AYVMAGVSPVSRLLGILCTSTVOLVTQMLPVCGLLDHVRENRGRIGSDLLNWCQIA 836
Db 771 AYVMAGVSPVSRLLGILCTSTVOLVTQMLPVCGLLDHVRENRGRIGSDLLNWCQIA 830
Qy 896 KMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLIDIDETEHADGKVPKIKMA 896
Db 831 KMSYLEDLVLRDLAARNVLKSPNHVKITDFGLARLLIDIDETEHADGKVPKIKMA 890
Qy 956 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
Db 891 LESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
Qy 1016 YVIMVWKWMDSCRRFRFRELVSERFMRADPQRFVYVIONEDLGASPLDSTFYRSILLE 1016
Db 951 YVIMVWKWMDSCRRFRFRELVSERFMRADPQRFVYVIONEDLGASPLDSTFYRSILLE 1010
Qy 1076 DDMGDLVDAEYLVLPQGGFCPPDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEA 1076
Db 1011 DDMGDLVDAEYLVLPQGGFCPPDPAPGAGMVHRRSSSTRSGGDLTLGLPSEEEA 1070
Qy 1136 PRSPLAPSEGAGSDVFDGDLGMGAAGLQSLPHTHDPSPLOQYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGDLGMGAAGLQSLPHTHDPSPLOQYSEDPTVPLPSETDGYVAP 1130
Qy 1196 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVVVKOVFAFGGA 1196
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAATLERAKTILSPGKGVVVKOVFAFGGA 1190
Qy 1256 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLG 1250
Qy 1261 LDVVPV 1261
Db 1251 LDVVPV 1255
```

## RESULT 13

AAW51143

ID AAW51143 standard; Protein; 1255 AA.

XX

AC AAW51143;

XX

17-JUN-2002 (first entry)

DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

XX

```
KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KW tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Domain 1..653
FT /note= "extracellular domain"
FT Domain 676..1255
FT /note= "intracellular domain"
FT Domain 990..1255
FT /note= "phosphorylation domain"
XX
PN WO200212341-A2.
XX
XX 14-FEB-2002.
XX
XX 03-AUG-2001; 2001WO-US24283.
XX
XX 03-AUG-2000; 2000US-0632507.
XX
XX (CORI-) CORIXA CORP.
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Cheever MA, Gheysen D;
XX
XX MPI; 2002-241743/29.
XX N-PSDB; ABA92250.
XX
```

Her-2/neu fusion protein for treating or preventing cancer by eliciting or enhancing an immune response to the protein, has Her-2/neu extracellular domain fused to Her-2/neu intracellular or phosphorylation domain -

Claim 68; Fig 7; 141pp; English.

The present sequence is that of human Her-2/neu (p185 glycoprotein or c-erbB2), an oncogenic self-protein and target for anti-cancer vaccines. The Her-2/neu gene is amplified and p185 is overexpressed in a variety of cancers, including breast, ovarian, colon, lung and prostate cancer. Her-2/neu is a member of the tyrosine kinase family of receptor-like glycoproteins. It comprises an extracellular domain with homology to the epidermal growth factor receptor (EGFR), a highly hydrophobic transmembrane domain and a C-terminal intracellular domain that also shows homology to EGFR. Its overexpression correlates with a poor prognosis in breast and ovarian cancers. The invention provides Her-2/neu fusion proteins, nucleic acids encoding them, viral vectors, and vaccines comprising the fusion proteins or nucleic acid molecules. In preferred fusion proteins, the extracellular domain of a Her-2/neu protein is fused to a Her-2/neu intracellular domain or phosphorylation domain (or its DeltaPD fragment). An immune response to Her-2/neu protein is elicited or enhanced by administering the fusion protein in the form of a vaccine, or by transfecting cells of an animal *ex vivo* with a nucleic acid encoding the fusion protein, and delivering the transfected cells to the animal. The fusion proteins, nucleic acids, and isolated specific T-cells are useful for inhibiting the development of a cancer, especially breast, ovarian, colon, lung or prostate cancer in a patient. T cells that specifically react with a Her-2/neu fusion protein can be used to remove tumour cells from a sample in order to inhibit the development of cancer in a patient.

Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 23; Length 1255;

Best Local Similarity 98.1%; Pred. No. 0;

Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

```
Qy 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRHLVGGCVVQGNL 60
Db 1 MELAALCRWGLLLALLPPGAASQVCTGTDMLRLPASPETHDMLRHLVGGCVVQGNL 60
```

QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
DB |||||  
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120  
DB |||||  
QY 121 DFLNNFNFTVFWLVRPKVSAS---HLEQURSLTEILKGGVLIQRNPQLCYQDTILWK 176  
DB |||||  
QY 121 DFLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQRNPQLCYQDTILWK 170  
DB |||||  
QY 177 DIFHKNNQALTLIDTNRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVCAGCARCKGP 236  
DB |||||  
QY 171 DIFHKNNQALTLIDTNRSRACHPCSPMCKGSRGWGESSEDCQSLTRTVCAGCARCKGP 230  
DB |||||  
QY 237 LPTDCHEOCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNBEGRYT 296  
DB |||||  
QY 231 LPTDCHEOCAAGCTGPKHSDCLACLHFNHSGICELHCPALVTYNTDTFESMPNBEGRYT 290  
DB |||||  
QY 297 FGASCVTACPNYLSLTVGSCSLVCPHLNHQEVTAEDGTQRCBKCKPCARVCYGLGMEHL 356  
DB |||||  
QY 291 FGASCVTACPNYLSLTVGSCSLVCPHLNHQEVTAEDGTQRCBKCKPCARVCYGLGMEHL 350  
DB |||||  
QY 357 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDPPASNTAPLOPEQLQVFTELEETGYL 416  
DB |||||  
QY 351 REVRAVTSANIOEFAGCKKIFGSLAPLPESFGDPPASNTAPLOPEQLQVFTELEETGYL 410  
DB |||||  
QY 417 YISAMPDLSLVDLVSFQNLQVIRGRILHNGAYSITLQGLGISWGLRLSRELGSGLAIHH 476  
DB |||||  
QY 411 YISAMPDLSLVDLVSFQNLQVIRGRILHNGAYSITLQGLGISWGLRLSRELGSGLAIHH 470  
DB |||||  
QY 477 NTHLCFVHTVPWDQLFRNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVN 536  
DB |||||  
QY 471 NTHLCFVHTVPWDQLFRNPQHALLHTANRPEDECVGEGGLACHQLCARGHCWGPPTQCVN 530  
DB |||||  
QY 537 CSQFLRGQCEVBEICRVLQGLPREYVNAHCLPCHPEQCQNGSVTCFGEADQCACAHY 596  
DB |||||  
QY 531 CSQFLRGQCEVBEICRVLQGLPREYVNAHCLPCHPEQCQNGSVTCFGEADQCACAHY 590  
DB |||||  
QY 597 KDPFPCVACRPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPORASP 656  
DB |||||  
QY 591 KDPFPCVACRPSGVKPDLSYMPIWKFPDEEGACQPCPINCTHSCVDLDDKGCAPORASP 650  
DB |||||  
QY 657 LTSIVSAVVGILLVVVLGVVFGILIKRROQKIRKYMTRRLLOTELVEPLTPSGAMPNOA 716  
DB |||||  
QY 651 LTSIVSAVVGILLVVVLGVVFGILIKRROQKIRKYMTRRLLOTELVEPLTPSGAMPNOA 710  
DB |||||  
QY 717 QMRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDE 776  
DB |||||  
QY 711 QMRILKETELRKVKVLGSGAFGVYKGIWIPGENVKIPVAIKVLRNTPSKANKEILDE 770  
DB |||||  
QY 777 AYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRLGSLQDILLNMQMIA 836  
DB |||||  
QY 771 AYVMAGVSPYVSRLLGICLTSTVOLVTOLMPYGCLLDHVRENRLGSLQDILLNMQMIA 830  
DB |||||  
QY 837 KGMSYLEVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGGKVPKMA 896  
DB |||||  
QY 831 KGMSYLEVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGGKVPKMA 890  
DB |||||  
QY 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 956  
DB |||||  
QY 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 950  
DB |||||  
QY 957 VYIMVVKCWMIDSECRPRFRELVSFSPRWARDPQRVVIQNEIDLGPASPLDSTFYRSLLE 1016  
DB |||||  
QY 951 VYIMVVKCWMIDSECRPRFRELVSFSPRWARDPQRVVIQNEIDLGPASPLDSTFYRSLLE 1010  
DB |||||  
QY 1017 DDDMGDLVDAEYLVPOQGFCCFPDPAAGGGMVHRHRSSTSRGGDLTLGLEPSEEA 1076  
DB |||||  
QY 1011 DDDMGDLVDAEYLVPOQGFCCFPDPAAGGGMVHRHRSSTSRGGDLTLGLEPSEEA 1070  
DB |||||  
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLRTHDPSPLQRYSEDPTVPLPSETDGYVAP 1136  
DB |||||  
QY 1071 PRSPLAPSEGAGSDVFDGLGMAAKGLQSLRTHDPSPLQRYSEDPTVPLPSETDGYVAP 1130  
DB |||||  
QY 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSFGKNGVVKDVFAPGGA 1196

DB 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAGATLERAKTLSFGKNGVVKDVFAPGGA 1190  
QY 1197 VENPEYLTTPQGGAAPOPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256  
DB 1191 VENPEYLTTPQGGAAPOPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250  
QY 1257 LDVPV 1261  
DB 1251 LDVPV 1255

RESULT 14  
AAU77114  
ID AAU77114 standard; Protein; 1255 AA.  
XX  
AC AAU77114;  
XX  
DT 05-JUN-2002 (first entry)  
XX  
DE Human Her-2/neu polypeptide.  
XX  
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;  
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;  
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;  
KW Hodgkin's lymphoma; T cell therapy.  
XX  
OS Homo sapiens.  
XX  
PN WO200213847-A2.  
XX  
PD 21-FEB-2002.  
XX  
PF 13-AUG-2001; 2001WO-US25408.  
XX  
PR 14-AUG-2000; 2000US-0638280.  
XX  
PR 28-SEP-2000; 2000US-0675904.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Gaiger A, Cheever MA, Hand-zimmermann S;  
DR WPI; 2002-280741/32.  
DR N-PSDB; ABK10730.  
XX  
PT Inhibiting haematological malignancy development by administering  
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide  
PT encoding the polypeptide, or antigen presenting cells expressing the  
PT polypeptide -  
XX  
PS Disclosure; Page 71-74; 74pp; English.  
XX  
CC The invention relates to a method for inhibiting development of  
CC haematological malignancy in a patient by administering a polypeptide  
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
CC encoding the polypeptide. Antigen presenting cells that express the  
CC protein can also be administered. The sequences are used for inhibiting  
CC development of haematological malignancy such as acute myelogenous  
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
XX  
SQ Sequence 1255 AA;

Query Match 98.0%; Score 6708; DB 23; Length 1255;  
Best Local Similarity 98.1%; Pred. No. 0;  
Matches 1241; Conservative 1; Mismatches 9; Indels 14; Gaps 2;

QY 1 MELAAICRWGLLALLPPCAASTQVCTGTDMLRLPASPEHLDMRLHYQCGVQGNL 60  
DB 1 MELAAICRWGLLALLPPCAASTQVCTGTDMLRLPASPEHLDMRLHYQCGVQGNL 60  
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIHNOVROVPLQRLRIVRGTQLFEDNYALAVLDNG 120

```
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNFNTVSWLRVPKVSAS-----HLEQLSLTEILKGGVLIQORNPOLCYQDITLWK 176
Db 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWK 170
Qy 177 DIFHKKNQALTLIDTNRSRACHPCSPMCKSGCWSESSEDCOSLTRTVCAGGCARCKGP 236
Db 171 DIFHKKNQALTLIDTNRSRACHPCSPMCKSGCWSESSEDCOSLTRTVCAGGCARCKGP 230
Qy 237 LPTDCHECAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 296
Db 231 LPTDCHECAAGCTGPKHSDCLACLFHFNHSGICELHCPALVTYNTDTFESMPNPEGRYT 290
Qy 297 FGASCVTACPYNYLSLTVGSCSLVCPHNOEVTAEQGTQCEKSKPCARVCYGLGMEHL 356
Db 291 FGASCVTACPYNYLSLTVGSCSLVCPHNOEVTAEQGTQCEKSKPCARVCYGLGMEHL 350
Qy 357 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 416
Db 351 REVRAVTSANIOEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 410
Qy 417 YISAWPDSLPDLVSFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRSLRELGLALIIH 476
Db 411 YISAWPDSLPDLVSFQNLQVIRGRIILHNGAYSILTQGLGISWLGSLRSLRELGLALIIH 470
Qy 477 NTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPQTQCVN 536
Db 471 NTHLCFVHTVPMQDLFRNPHQALLHTANRPEDECVGEGLACHOLCARGHCWGPQTQCVN 530
Qy 537 CSOFLRQCEVEECRVLQGLPREYVNAHCLCPHCECPONGSVTCFGEADOCVACAHY 596
Db 531 CSOFLRQCEVEECRVLQGLPREYVNAHCLCPHCECPONGSVTCFGEADOCVACAHY 590
Qy 597 KDPFFCVARCPGVKPDLSYMPYIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEQRAS 656
Db 591 KDPFFCVARCPGVKPDLSYMPYIWKFPDEGACQPCPINCTHSCVDLDDKGCFAEQRAS 650
Qy 657 LTSIVSAVGIILVAVLVGVLGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQA 716
Db 651 LTSIVSAVGIILVAVLVGVLGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQA 710
Qy 717 QMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKILDE 776
Db 711 QMRILKETELRKVKVLSGAGFTVYKGIWIPDGENVKIPVAIKVIRENTSPKANKILDE 770
Qy 777 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTLQMPYGCILLDHVRENRLGSDLLNWCWQIA 836
Db 771 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTLQMPYGCILLDHVRENRLGSDLLNWCWQIA 830
Qy 837 KGMSTYEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPWKMA 896
Db 831 KGMSTYEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETEHADGGKVPWKMA 890
Qy 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPPOPICTID 950
Qy 957 VYMIWKWMIIDSECPRRRELVSFPRMARDQRFVVIQNEDLGPASPLDSTFYRSLLE 1016
Db 951 VYMIWKWMIIDSECPRRRELVSFPRMARDQRFVVIQNEDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVPOQGFCCPDPAAGAGWVHRRSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMGAAGLQSLPHTDPSPLQRYSEDPTVPLPSETDGYVAP 1130
Qy 1137 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAATLERAKTUSPKNGVGVKDVFAFGGA 1196
Db 1131 LTCSPQEVYNQDVRPQPPSPREGPLPAARPAATLERAKTUSPKNGVGVKDVFAFGGA 1190
Qy 1197 VENPEYLTTPQGGNAOPHPHPAFSPAFDNLTYWDQDPPPERGAPSTFKGTTPTAENPEYLG 1256
Db 1191 VENPEYLTTPQGGNAOPHPHPAFSPAFDNLTYWDQDPPPERGAPSTFKGTTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255
```

## RESULT 15

```
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN WO9316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU ) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
XX
WI: 1993-272889/34.
XX
N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;
```

```
Query Match 97.3%; Score 6665; DB 14; Length 1433;
Best Local Similarity 97.5%; Pred. No. 0;
Matches 1234; Conservative 3; Mismatches 14; Indels 14; Gaps 2;
```

```
Qy 1 MELAALCRVGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLHYQGCVQVGNL 60
Db 1 MELAALCRVGLLALLPPGAASQVCTGDMKRLPASPTHLDMLRHLHYQGCVQVGNL 60
Qy 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFLQDIQEVQGVLIHAHQVQVPLQRLRIVRGTLFEDNYALAVLDNG 120
Qy 121 DPLNNFNTVSWLRVPKVSAS-----HLEQLSLTEILKGGVLIQORNPOLCYQDITLWK 176
Db 121 DPLNN-----TTPVTGASPGGLRELQRLSLTEILKGGVLIQORNPOLCYQDITLWK 170
Qy 177 DIFHKKNQALTLIDTNRSRACHPCSPMCKSGCWSESSEDCOSLTRTVCAGGCARCKGP 236
Db 171 DIFHKKNQALTLIDTNRSRACHPCSPMCKSGCWSESSEDCOSLTRTVCAGGCARCKGP 230
```

Search completed: July 22, 2003, 08:40:33  
Job time : 43.1589 secs

```
Qy 237 LPTDCHEQACGCTGPKHSDCLACLUHFNHSGICELHCPALVTYNTDTFESMNPBGRYT 296
Db 231 LPTDCHEQACGCTGPKHSDCLACLUHFNHSGICELHCPALVTYNTDTFESMNPBGRYT 290
Qy 297 FGASCVTACPNYVNSTDVGSCTLVCPLNHNOEVTAEQDCEKCKSPCARVCYGLGMEHL 356
Db 291 FGASCVTACPNYVNSTDVGSCTLVCPLNHNOEVTAEQDCEKCKSPCARVCYGLGMEHL 350
Qy 357 REVRAVTSANIQEFACCKIIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEEITGYL 416
Db 351 REVRAVTSANIQEFACCKIIFGSLAFLPESFDGDPASNTAPLQPEHLQVFEITLQITGYL 410
Qy 417 YISAMPDPSLPDLVSFQNLQVIRGRILHNGAYSITLQGLGSMGLRSLRELGSGLALIIH 476
Db 411 YISAMPDPSLPDLVSFQNLQVIRGRILHNGAYSITLQGLGSMGLRSLRELGSGLALIIH 470
Qy 477 NTHLCFVHTVPWDOLFERNPHQALLHTANRPEDECVEGLACHOLCARGHGWGPGTQCVN 536
Db 471 NTHLSFVHTVPWDOLFERNPHQALLHTANRPEDECVEGLACHOLCARGHGWGPGTQCVN 530
Qy 537 CSQFLRGQECVECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFPGPEADQCACAHY 596
Db 531 CSQFLRGQECVECRVLQGLPREYVNAHCLPCHPECQPNQSVTCFPGPEADQCACAHY 590
Qy 597 KDPFFCVARCPGSGVKRDPDLSPYMPIWKPDEEGACQPCPINCTHSCVDLDDKGCPEORASP 656
Db 591 KDPFFCVARCPGSGVKRDPDLSPYMPIWKPDEEGACQPCPINCTHSCVDLDDKGCPEORASP 650
Qy 657 LTSISAVVGILLVVVLGVVFGILIKRQOKIRKKTMRRLQETELVEPLTPSGAMPNOA 716
Db 651 LTSISAVVGILLVVVLGVVFGILIKRQOKIRKKTMRRLQETELVEPLTPSGAMPNOA 710
Qy 717 QMRILKETELRKVKVLGSGAFGTGKGIWIPDGENYKIIPVAIKVRENTSPKANKELDE 776
Db 711 QMRILKETELRKVKVLGSGAFGTGKGIWIPDGENYKIIPVAIKVRENTSPKANKELDE 770
Qy 777 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTOLMPYGCLLDHVRENRRGLGSODLLNWCQIA 836
Db 771 AYVMAGVGSPPYVSRLLIGICLTSTVQLVTOLMPYGCLLDHVRENRRGLGSODLLNWCQIA 830
Qy 837 KGMVLYEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDETEYHADGGKVPKQMA 896
Db 831 KGMVLYEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDDETEYHADGGKVPKQMA 890
Qy 897 LESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 956
Db 891 LESILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICTID 950
Qy 957 VYMIWVKWMIIDSECRPRPRELVSEFSRWARDPQRFVJONEDLGPASPLDSTFYRSLLE 1016
Db 951 VYMIWVKWMIIDSECRPRPRELVSEFSRWARDPQRFVJONEDLGPASPLDSTFYRSLLE 1010
Qy 1017 DDDMGDLVDAEYLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEEA 1076
Db 1011 DDDMGDLVDAEYLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEEA 1070
Qy 1077 PRSPLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSLQRYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSPLAPSEGAGSDVFDGLGMAKGLQSLPHTDPSLQRYSEDPVPLPSETDGYVAP 1130
Qy 1137 LTCSPQPEYVNOQDVYRPPQPPSPREGPLPAARFAGATLERAKTSLSPKNGVWVDVFAFGA 1196
Db 1131 LTCSPQPEYVNOQDVYRPPQPPSPREGPLPAARFAGATLERAKTSLSPKNGVWVDVFAFGA 1190
Qy 1197 VENPEYLTPOGGAAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAAPQHPHPPAFSPAFDNLYYWDQDPPERGAPPSTFKGTPTAENPEYLG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:25:54 ; Search time 23.0157 Seconds  
(without alignments)  
5267.077 Million cell updates/sec

Title: SEQ4-149-163-14  
Perfect score: 6848  
Sequence: 1 MELAALCRWGLLLALLPPGA.....TFKGTPAENPEYGLDVPV 1261  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues  
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 73:.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	6709	98.0	1255	1	A24571	protein-tyrosine k
2	5918	86.4	1260	1	TVRNU	protein-tyrosine k
3	5914.5	86.4	1254	2	I48161	p-185 precursor -
4	3164	46.2	1210	1	GOHUE	epidermal growth f
5	3141	45.9	1210	2	A53183	epidermal growth f
6	3110.5	45.4	1223	1	TVCHLV	epidermal growth f
7	2972.5	43.4	1308	2	A47253	epidermal growth f
8	2674	39.0	1166	1	S06142	protein-tyrosine k
9	2418.5	35.3	1342	2	A36223	kinase-related tra
10	2341.5	34.2	1339	2	JC4387	epidermal growth f
11	1766.5	25.8	698	1	TVFVLV	protein-tyrosine k
12	1703	24.9	604	1	TVYUHV	protein-tyrosine k
13	1647	24.1	544	2	S35745	epidermal growth f
14	1642.5	24.0	1330	1	GQFFE	kinase-related tra
15	1640	23.9	545	2	S00727	protein-tyrosine k
16	1623	23.7	540	2	B44776	protein-tyrosine k
17	1621	23.7	540	1	TVFVEB	epidermal growth f
18	1531	22.4	644	2	A36325	protein-tyrosine k
19	1302	19.0	1323	2	E88257	epidermal growth f
20	1302	19.0	1374	2	S70712	protein let-23 [im
21	1208	17.6	1369	2	S70713	protein-tyrosine k
22	1157	16.9	1717	1	A45558	epidermal growth f
23	1142	16.7	527	2	A42032	epidermal growth f
24	987.5	14.4	843	2	A27131	epidermal growth f
25	806.5	11.8	346	2	S13807	protein-tyrosine k
26	754.5	11.0	311	2	S13808	protein-tyrosine k
27	734	10.7	1363	2	T43220	insulin-like growth
28	704	10.3	1382	1	INHUR	insulin receptor p
29	698.5	10.2	1300	2	A36502	insulin receptor-r

## RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein e

C:Species: Homo sapiens (man)

C:Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C:Accession: A24571; A24591; A44188; B44188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T

Nature 319, 230-234, 1986

A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 &lt;YAM&gt;

A:Cross-references: GB:X03363; NID:g31197; PIDN:CAA27060.1; PID:g31198

R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 &lt;SEM&gt;

A:Cross-references: GB:M11767; NID:g182163; PIDN:AAA35808.1; PID:g553282

R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg

Science 230, 1132-1139, 1985

A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro

A:Reference number: A44188; MUID:86070181; PMID:2999374

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 &lt;COU1&gt;

A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALL', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 &lt;COU2&gt;

A:Cross-references: GB:M11730; NID:g183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 832-909 &lt;REX&gt;

A:Cross-references: GB:L29395; NID:g459807; PIDN:AAA35809.1; PID:g459808

R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A:Status: translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-191 &lt;TAL&gt;

A:Cross-references: GB:M16792; NID:q183983; PIDN:AAA58637.1; PID:q553332  
C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
C:Genetics:  
A:Gene: GDB:ERBB2; NGL; NEU; HER-2  
A:Cross-references: GDB:120613; OMIM:164870  
A:Map position: 17q21.1-17q21.1  
A:Introns: 25/1; 75/3; 147/1; 883/3  
A:Note: the list of introns is incomplete  
C:Function:  
A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphotyrosine  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
F:22-653/Domain: extracellular #status predicted <EXT>  
F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>  
F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
F:654-675/Domain: transmembrane #status predicted <TM>  
F:676-1255/Domain: intracellular #status predicted <INT>  
F:718-983/Domain: protein kinase homology <KIN>  
F:726-734/Region: protein kinase ATP-binding motif  
F:68,124,187,259,530,571,629/Binding site: carboxydrate (Asn) (covalent) #status predicted  
F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:753/Active site: Lys #status predicted  
F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 98.0%; Score 6709; DB 1; Length 1255;  
Best Local Similarity 98.0%; Pred. No. 2.9e-266;  
Matches 1240; Conservative 7; Mismatches 4; Indels 14; Gaps 3;

QY 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRLPASPTHLDMLRLHYQSCVQVQGNL 60  
DB 1 MELAALCRWGLLLALLPGCAASTQVCTGDMKRLRLPASPTHLDMLRLHYQSCVQVQGNL 60

QY 61 EUTYLPNTASLSFLQDIOEQVGVLIQNRQVPLQRLRIRVGTQLFEDNVALAVLDNG 120  
DB 61 EUTYLPNTASLSFLQDIOEQVGVLIQNRQVPLQRLRIRVGTQLFEDNVALAVLDNG 120

QY 121 DPLNNTPTVTGASPGSLRLQLRLSLEILKGGVLIQNRQVPLQRLRIRVGTQLFEDNVALAVLDNG 177  
DB 121 DPLNNTPTVTGASPGSLRLQLRLSLEILKGGVLIQNRQVPLQRLRIRVGTQLFEDNVALAVLDNG 177

QY 178 SEWLRVPKVSASHLE-NRGRACHPCSPMCKSRGCMGSESDCQSLTRTVCAAGCARCKGP 236  
DB 178 -----QALTLIDNRSRACHPCSPMCKSRGCMGSESDCQSLTRTVCAAGCARCKGP 230

QY 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPEGRYT 296  
DB 237 LPTDCHEQCAAGCTGPKHSDCLACLFHNSHIGICELHCPALVTYNTDTFESMPNPEGRYT 290

QY 297 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEQDTCRCKSPCARVCYGLGMEHL 356  
DB 297 FGASCVTACPNYLSLTDVGSCTLVCPHNOEVTAEQDTCRCKSPCARVCYGLGMEHL 350

QY 357 REVRAVTSANIOEFACCKIFGSLAPLPSFDGDPASNTAPLOEQVLFTELEITGYL 416  
DB 357 REVRAVTSANIOEFACCKIFGSLAPLPSFDGDPASNTAPLOEQVLFTELEITGYL 410

QY 417 YISAMPDLSPLDLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRLSRELGSGLALIH 476  
DB 417 YISAMPDLSPLDLSVFQNLQVIRGRILHNGAYSLTLQGLGSLWGLRLSRELGSGLALIH 470

QY 477 NTHLCFVHTVPDOLFRNPHQALLHTANRPECEVGEGLACHOLCARGHCWGPGTQCVN 536  
DB 477 NTHLCFVHTVPDOLFRNPHQALLHTANRPECEVGEGLACHOLCARGHCWGPGTQCVN 530

QY 537 CSQFLRGQCEVBECEVRLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHY 596  
DB 537 CSQFLRGQCEVBECEVRLQGLPREYNARHCLPCHPECPQNGSVTCFGEADQCVACAHY 590

QY 597 KDPPFCVAPCSGVKPDLSYMPITWKPDEBAGACQPCPINCTHSCVDLDDKGCPAEQRAS 656

Db 591 KDPPFCVAPCSGVKPDLSYMPITWKPDEBAGACQPCPINCTHSCVDLDDKGCPAEQRAS 650  
QY 657 LTSIVSAVVGILLVVLGVVFGILLKRRQOKIRKYMRLLOETELVELPLTSGAMPNOA 716  
Db 651 LTSIIISAVVGILLVVLGVVFGILLKRRQOKIRKYMRLLOETELVELPLTSGAMPNOA 710  
QY 717 QMRILKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 776  
Db 711 QMRILKETELRKVKVLGSGAGCTVYKGIWIPDGENVKIPVAIKVLRNTPSKANKEILDE 770  
QY 777 AYVMAGVSPVYSRLGICLTSTVOLVTQLMPYGCLLDHHVNRGRGLSGQDLLNMCQIA 836  
Db 771 AYVMAGVSPVYSRLGICLTSTVOLVTQLMPYGCLLDHHVNRGRGLSGQDLLNMCQIA 830  
QY 837 KCMSYLEVDVLRVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 896  
Db 831 KCMSYLEVDVLRVHRLDAAARNVLKSPNHVKITDFGLARLLDIDETEHADGKVPKMA 890  
QY 897 LESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICTID 956  
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFCAPYDGIIPAREIPDLLEKGERLPQPPICTID 950  
QY 957 VYIMVVKCWMIDSECRPRFRELVSFMRARDPQRFVWIONEDLGPASPLDSTFYRSLLE 1016  
Db 951 VYIMVVKCWMIDSECRPRFRELVSFMRARDPQRFVWIONEDLGPASPLDSTFYRSLLE 1010  
QY 1017 DDMGDLVDAEYLYVPOQGFCCPDAPGAGGVMHHRHSSSTRSGGDLTLGLEPSEEA 1076  
Db 1011 DDMGDLVDAEYLYVPOQGFCCPDAPGAGGVMHHRHSSSTRSGGDLTLGLEPSEEA 1070  
QY 1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLTPHPSPLQRYSEDPVLPSETDGYVAP 1136  
Db 1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLTPHPSPLQRYSEDPVLPSETDGYVAP 1130  
QY 1137 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERAKTSLSPGKNGVKDVPFAGGA 1196  
Db 1131 LTCSPQPEYVNPQDVRPQPPSPREGPLPAARPAAGATLERPKTSLSPGKNGVKDVPFAGGA 1190  
QY 1197 VENPEYLYPQGGAAQPPHPPAFSAFONLYYWDQDPPRGAPPSTFKGTPTAENPEYLG 1256  
Db 1191 VENPEYLYPQGGAAQPPHPPAFSAFONLYYWDQDPPRGAPPSTFKGTPTAENPEYLG 1250  
QY 1257 LDVPV 1261  
Db 1251 LDVPV 1255

RESULT 2  
TVRTNU  
protein-tyrosine kinase (EC 2.7.1.112) new precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999  
C:Accession: A24562; A61204  
R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.  
Nature 319, 226-230, 1986  
A:Title: The new oncogene encodes an epidermal growth factor receptor-related protein  
A:Reference number: A24562; MUID:86118662; PMID:3945311  
A:Accession: A24562  
A:Molecule type: mRNA  
A:Residues: 1-1260 <BAR>  
A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:q56746  
R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cohen  
Carcinogenesis 12, 1975-1978, 1991  
A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals a  
2-thiazoylformamide or N-methyl-N-nitrosourea.  
A:Reference number: A61204; MUID:92035293; PMID:1682063  
A:Accession: A61204  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 637-663, 'V', 665-702 <MAS>  
A:Note: authors translated the codon GCA for residue 25 as Val  
C:Genetics:  
A:Gene: neu